

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 23:15:18 ; Search time 1854.01 seconds
(without alignments)
11362.978 Million cell updates/sec

Title: US-09-645-192-1
Perfect score: 1362
Sequence: 1 atgaagatattcaatgtta.....atctactaccacatcatga 1362

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_da1:*
- 2: gb_da2:*
- 3: gb_da3:*
- 4: gb_da4:*
- 5: gb_da5:*
- 6: gb_da6:*
- 7: gb_da7:*
- 8: gb_da8:*
- 9: gb_da9:*
- 10: gb_da10:*
- 11: gb_da11:*
- 12: gb_da12:*
- 13: gb_da13:*
- 14: gb_da14:*
- 15: gb_da15:*
- 16: gb_da16:*
- 17: gb_da17:*
- 18: gb_da18:*
- 19: gb_da19:*
- 20: gb_da20:*
- 21: gb_da21:*
- 22: gb_da22:*
- 23: gb_da23:*
- 24: gb_da24:*
- 25: gb_da25:*
- 26: gb_da26:*
- 27: gb_da27:*
- 28: gb_da28:*
- 29: gb_da29:*
- 30: gb_da30:*
- 31: gb_da31:*
- 32: gb_da32:*
- 33: gb_da33:*
- 34: gb_da34:*
- 35: gb_da35:*
- 36: gb_da36:*
- 37: gb_da37:*
- 38: gb_da38:*
- 39: gb_da39:*
- 40: gb_da40:*
- 41: gb_da41:*
- 42: gb_da42:*
- 43: gb_da43:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vl:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vl1:*

59: gb_vl2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_pr10:*

95: gb_pr11:*

96: gb_pr12:*

97: gb_pr13:*

98: gb_pr14:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	10	AX087935
2	1362	100.0	3435	89	AF132035
3	207.2	15.2	3508	94	AB037596
4	194.8	14.3	108873	58	AF318573
5	194.4	14.3	1206	94	AB037597
6	191.8	14.1	1317	10	AX087948
7	191.8	14.1	1317	88	AF038650
8	191.8	14.1	2128	88	AF102542

9	191.8	14.1	2216	9	AX045249	Sequence
10	191.8	14.1	181983	57	AC022480	Homo sapi
11	191.6	14.1	2017	58	AF231105	Bovine he
12	191.2	14.0	2393	94	MM068182	U68182 Mus musculu
13	187.4	13.8	87636	90	AL138039	AL138039 Human DNA
14	187.4	13.8	176544	78	AL138877	AL138877 Homo sapi
15	187.4	13.8	196788	66	AC021626	AC021626 Homo sapi
16	185.2	13.6	1203	10	AX087950	AX087950 Sequence
17	185.2	13.6	1807	10	I16989	I16989 Sequence 13
18	185.2	13.6	1807	10	I16989	I16989 Sequence 13
19	185.2	13.6	1807	93	HSBNA8T	HSBNA8T H. sapiens m
20	185.2	13.6	1807	97	HBMN8AT	HBMN8AT Human I bet
21	184.2	13.5	157102	75	AC076970	AC076970 Homo sapi
22	183.6	13.5	155359	90	AL358777	AL358777 Human DNA
23	183.6	13.5	196788	66	AC021626	AC021626 Homo sapi
24	181.2	13.3	252504	75	AC073815	AC073815 Mus muscu
25	180.8	13.3	2217	9	AX045251	AX045251 Sequence
26	180.8	13.3	2217	9	AX045251	AX045251 Sequence
27	179.6	13.2	2033	94	MM019965	MM019965 Mus musculu
28	179.6	13.2	2033	94	MM019965	MM019965 Mus musculu
29	175.2	12.9	183858	90	AL161626	AL161626 Human DNA
30	174.4	12.8	5010	95	S79797	S79797 enzymatic g
31	174	12.8	2060	93	HSIGN1	HSIGN1 Human beta-
32	173.2	12.7	2023	94	DB7332	DB7332 Mus musculu
33	173.2	12.7	2047	94	DB7332	DB7332 Mus musculu
34	172.8	12.7	4179	58	BM44R6F	BM44R6F Bovine hearp
35	172	12.6	1287	10	AX087946	AX087946 Sequence
36	172	12.6	1287	10	AX087946	AX087946 Sequence
37	172	12.6	2105	10	I16140	I16140 Sequence 3
38	172	12.6	2105	10	I16140	I16140 Sequence 3
39	172	12.6	2105	10	I16140	I16140 Sequence 3
40	172	12.6	2105	10	I16140	I16140 Sequence 3
41	164.8	12.1	2435	97	HM0C2CNT	HM0C2CNT Homo sapien
42	158.4	11.6	94171	93	HS1153D9	HS1153D9 Bos taurus
43	121.2	8.9	378	10	I16983	I16983 Sequence 1
44	121.2	8.9	378	10	I16983	I16983 Sequence 1
45	113.4	8.3	63853	77	AC087663	AC087663 Homo sapi

ALIGNMENTS

RESULT 1
 LOCUS AX087935 1362 bp DNA PAT 17-MAR-2001
 DEFINITION Sequence 1 from Patent WO0114535.
 ACCESSION AX087935
 VERSION AX087935.1 GI:13396913

KEYWORDS

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

Udp-n-acetylglucosamine: galactose-4-epimerase (NADP+ dependent)
 -acetylglucosamine-4-epimerase (NADP+ dependent)
 Patent: WO 0114535-A 1 01-MAR-2001;
 Schwientek, T. and Clausen, H.
 Schwenk, T. and Clausen, H. (1999) J. Biol. Chem. 274: 1153-1159

JOURNAL

FEATURES

SOURCE

CDS

1..1362
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..1362
 /note="The gene of C2GNT3"
 /codon_start=1
 /protein_id="CAC34674.1"
 /db_xref="GI:13396913"
 /translation="MTFKCYFKHTLOQKVFILPTIMLSLKLVNRLPQKDI
 LVEISTSPFVRNRTTHVDVRYKNSGIYEOPLTSGSLRRLIDLEDD
 VVAMTSDCDIYQIRGAKLVSKESFPIALSLYHKAIVELIRAHYIQHNIY
 CIHYDRAPDITFKYAMNNLAKFSNFIASKLEAVEYAHISRLQADLNCUSDLSKSI

Query Match 100.0%; Score 1362; DB 10; Length 1362;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT
 410 a 260 c 279 g 413 t

1 atgaagatattcaatgtttatcttaaacatcccttaacagcaagttcatctgtt 60
 1 ATGAGGATTTCAATGTTTAAACATACCTTACGCAAGATTTTCACTCTGTT 60
 61 ttaacccatgctgctctctctttaaagcttctaagtgtgagcagctcttcgcaa 120
 61 TTAACCCATGCTGCTCTCTCTTTAAAGCTTCTAAATGAGACAGACTTCCGCAA 120
 121 aaagcaattctctgtgtgagtaaccttaagctctctgttgaagaagaatgac 180
 121 AAAGCAATTCTCTGTGTGAGTAACCTTACCTGCTGCTGTTGTAAGAACAGATAC 180
 181 actcatgttaagagatgaagtcagatgaagtcagatgttcggtatctatgaacagag 240
 181 ACTCATGTTAAGAGATGAAGTCAGATGAAGTCAGTAACTTGGGATCTATGAACAGAG 240
 241 cctttggaatttggaagaagtcgtgaataaagaagaagacatctatgttgagagat 300
 241 CCTTTGGAATTGGAAGAAGTCGTGAATAAAGAAGAAGACATCTATGTGAGAGAT 300
 301 gatgatgttggtgacatgacagatgacatgacatctatctcagactcgaagatgact 360
 301 GATGATGTGTGTCGACATGACCGATGATGTGACATTTATCAGACTCAGAGGTTATGCT 360
 361 caaagctgtgtcctcaagagagagaaagcttcccaatagcctattctgtgtgcac 420
 361 CAAAGCTGTGTCTCAAGAGAGAGAAAGCTTCCCAATAGCCTATTCTTGTGTGCAC 420
 421 aagaatgacattatggttgaagaagcttcaatgacatgacatgacatgacatgac 480
 421 AAGAATGACATTATGTTGAAGAAGCTTCAATGACATGACATGACATGACATGACAT 480
 481 tgcattccattatgacgtgaagacacatgacatgacatgacatgacatgacatgac 540
 481 TGCATTCATTATGACGTGAAGACACATGACATGACATGACATGACATGACATGACAT 540
 541 aagtgcttcccaatttcaatgacatgacatgacatgacatgacatgacatgacatgac 600
 541 AAGTGCTTCCCAATTTCATGACATGACATGACATGACATGACATGACATGACATGACAT 600
 601 tccagactccagctgatttaattgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660
 601 TCCAGACTCCAGCTGATTTAATTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
 661 aaatagttataactgctgt 720
 661 AAATAGTTATTAACCTGT 720
 721 tcagagtgaataaatacctcaatgagcaaatatgttggagacggtgaaaccccaaacagt 780
 721 TCAGAGTGAATAAATACTCAATGAGCAAAATATGTTGAGACAGGTAAACCCCAACAGT 780
 781 aaattggaagaatgacatgacatgacatgacatgacatgacatgacatgacatgacatgac 840
 781 AAATTGGAAGAATGACATGACATGACATGACATGACATGACATGACATGACATGACATGACAT 840
 841 ctaccataaggaacaacatctccaaaggaagacccccataacattcagatttgtt 900
 841 CTACCATAGGACAACATCTCCAAAGGAAGACCCCCATACATTAGATTGTTT 900

QY 901 ggcagtgcttatttcttgaagcaagcattgttaataatatattcaaacactccatc 960
 |||||
 Db 901 ggcagtgcttatttcttgaagcaagcattgttaataatatattcaaacactccatc 960
 |||||
 QY 961 gttaagaacttttctgctgtctaaagacacatactctcctgatgagcacttttgact 1020
 |||||
 Db 961 gttaagaacttttctgctgtctaaagacacatactctcctgatgagcacttttgact 1020
 |||||
 QY 1021 accttgatcgggttccagaataaccctggggagattccagaatcccccagatgctc 1080
 |||||
 Db 1021 accttgatcgggttccagaataaccctggggagattccagaatcccccagatgctc 1080
 |||||
 QY 1081 gatctgagagtaagactcgcctgttcaagtgaattacatgaaggcttttctatccc 1140
 |||||
 Db 1081 gatctgagagtaagactcgcctgttcaagtgaattacatgaaggcttttctatccc 1140
 |||||
 QY 1141 agtctgactggtatcacccttcgaagcgtgtgtatttaaggagctgcagaataaagttg 1200
 |||||
 Db 1141 agtctgactggtatcacccttcgaagcgtgtgtatttaaggagctgcagaataaagttg 1200
 |||||
 QY 1201 ctatcaagaatgacattgttctgttaataatttgaattcgaagtgagccctatctg 1260
 |||||
 Db 1201 ctatcaagaatgacattgttctgttaataatttgaattcgaagtgagccctatctg 1260
 |||||
 QY 1261 attaaatgcttgcagaagaacttgaagaacagcagagagacttgatcacttgcctca 1320
 |||||
 Db 1261 attaaatgcttgcagaagaacttgaagaacagcagagagacttgatcacttgcctca 1320
 |||||
 QY 1321 gaaaagttatttattgataagaatctcctaccatcatca 1362
 |||||
 Db 1321 gaaaagttatttattgataagaatctcctaccatcatca 1362
 |||||

RESULT 2

AF132035 3435 bp mRNA PRI 08-APR-2000
 LOCUS Homo sapiens core 2 beta-1,6-N-acetylglucosaminyltransferase 3
 DEFINITION (C2Gnt3) mRNA, complete cds.

ACCESSION AF132035
 VERSION AF132035.1 GI:7527463

KEYWORDS

SOURCE

ORGANISM

human.
 Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 3435)
 Schiøntek, T., Yeh, J.-C., Levery, S.B., Keck, B., Merckx, G., van

Kessel, A.G., Fukuda, M. and Clausen, H.
 Control of O-glycan branch formation. Molecular cloning and

characterization of a novel thymus-associated core 2 beta1,
 6-N-acetylglucosaminyltransferase

J. Biol. Chem. 275 (15), 11106-11113 (2000)

JOURNAL

MEDLINE

2 (bases 1 to 3435)
 Schiøntek, T. and Clausen, H.

Direct Submission
 Submitted (27-FEB-1999) School of Dentistry, University of

TITLE

JOURNAL

FEATURES

source

- gene

CDS.

/db_xref="taxon:9606"
 1.3435
 /gene="C2Gnt3"
 862..2223
 /gene="C2Gnt3"
 /codon_start=1
 /product=core 2 beta-1,6-N-acetylglucosaminyltransferase
 3
 /protein_id="AF63156.1"
 /db_xref="GI:7527464"
 /translation="MKIFKCYFHKTLQCKVFLFLTLMLSLIKLNNRLEPQKDIY
 LVEYSLSLSPFVNNRTYHWKDEVRYEYNSGIYEDPELIGSLKLRDIDLEDDD

VVAMTSDCDIYQTLRGYAOKLVSKSEKSFPIAVSLVYHKDAINVERLIIHYQNHIT
 CIHYDRKAPDQFFKVAAMNMLAKCFENFIASLKEAVEVAHSLROADLNCJSLDKSI
 QMKVYINCGODPDKSNFELVSEKLENGNMLETVKPPSKLEHRLRYDY
 EYVLTPTNTNKRPAHNIQIFVSGAVYISQAFVYIEFNSIVQDFPMASKOTYSP
 DEHWRATLIRYPIGELSRSAQDVSQDOSTRLVKNMYEGEFYPCGSHLSRVCI
 YGAELRWLIDGHWFAKFDPSKDPILIKCLAKLEBQCDWITLPSKLFMRNLT
 TTS"
 BASE COUNT 1032 a 678 c 651 g 1074 t
 ORIGIN

Query Match 100.0%; Score 1362; DB 89; Length 3435;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatttcaaatgttattttaaactaaccttcaagcagaagaatttcaatcctgttc 60
 |||||
 Db 862 atgaagatttcaaatgttattttaaactaaccttcaagcagaagaatttcaatcctgttc 921
 |||||
 QY 61 ttaacctatggtcgtcctcttctttaaagcttcaaatgtgagagactcttcgcaa 120
 |||||
 Db 922 ttaacctatggtcgtcctcttctttaaagcttcaaatgtgagagactcttcgcaa 981
 |||||
 QY 121 aagaacatttacttggcttgaagctaccccaagactcgccttcttgaagaacagatac 180
 |||||
 Db 982 aagaacatttacttggcttgaagctaccccaagactcgccttcttgaagaacagatac 1041
 |||||
 QY 181 actcatgtttaagaatgaatgaagtaagaatgaagtaagaatgaagtaagaatgaagta 240
 |||||
 Db 1042 actcatgtttaagaatgaatgaagtaagaatgaagtaagaatgaagtaagaatgaagta 1101
 |||||
 QY 241 ccttctgaaatttgaaagagctcgtgaagaataagaagaagagacatcacttgaatgagat 300
 |||||
 Db 1102 ccttctgaaatttgaaagagctcgtgaagaataagaagaagagacatcacttgaatgagat 1161
 |||||
 QY 301 gatgatcttctggaacatgaacagtgatcttgacattatcaagactcgaaggttactgct 360
 |||||
 Db 1162 gatgatcttctggaacatgaacagtgatcttgacattatcaagactcgaaggttactgct 1221
 |||||
 QY 361 caaagcttctcctaagaagagagaaagcttcccaatagcctattcttctgtgtccac 420
 |||||
 Db 1222 caaagcttctcctaagaagagagaaagcttcccaatagcctattcttctgtgtgtccac 1281
 |||||
 QY 421 aaagatcgaattatgcttgaagaagcttaccatgctatatacaacagcacaatatttac 480
 |||||
 Db 1282 aaagatcgaattatgcttgaagaagcttaccatgctatatacaacagcacaatatttac 1341
 |||||
 QY 481 tgcattcatttgcgtgaagcaccctgataaccttcaaaagttggcagaaacatttagct 540
 |||||
 Db 1342 tgcattcatttgcgtgaagcaccctgataaccttcaaaagttggcagaaacatttagct 1401
 |||||
 QY 541 aagtgctctccaatatttccattgcttcccaattagaagctgtggaatatgcccacatt 600
 |||||
 Db 1402 aagtgctctccaatatttccattgcttcccaattagaagctgtggaatatgcccacatt 1461
 |||||
 QY 601 tccagactcgaagctgattttaaattgcttgcggaacttccgaaggtttcaatccagttg 660
 |||||
 Db 1462 tccagactcgaagctgattttaaattgcttgcggaacttccgaaggtttcaatccagttg 1521
 |||||
 QY 661 aaatgtttatcaacttctgtggaagaatttccctcgtgaagtcgaatatttgaattgtgtg 720
 |||||
 Db 1522 aaatgtttatcaacttctgtggaagaatttccctcgtgaagtcgaatatttgaattgtgtg 1581
 |||||
 QY 721 tcaaggttgaaaaacccaatgagagcaaatgtttggaagaggtggaaccccccagaaggt 780
 |||||
 Db 1582 tcaaggttgaaaaacccaatgagagcaaatgtttggaagaggtggaaccccccagaaggt 1641
 |||||
 QY 781 aaattggaagaattcacttaccatcatgaacttgaagcgtgtgacttgaatlatgtgaag 840
 |||||
 Db 1642 aaattggaagaattcacttaccatcatgaacttgaagcgtgtgacttgaatlatgtgaag 1701
 |||||
 QY 841 ctaccaataagaacaacatctccaaggaagacccccccataaatctcaatatttctt 900
 |||||

Db 1702 CTACCAATTAAGGACCAATCTCCAGGAAGACACCCCCATTAACATTCAGATATTGTT 1761

Qy 901 ggagctctatttctgtttaagcaagcattgtttaatatatttccaaacatccatc 960

Db 1762 GGCAGCTCTATTGTTGTTTAAAGCAATTTGTTAAATATTATTTCACAACTCATC 1821

Qy 961 gtcaagaacttttctgctgtcctaaagacacatactctctctgatgacactttggct 1020

Db 1822 GTTCAAGACTTTTGGCTGGCTTAAAGACACATACCTCTGATGAGACTTTGGGCT 1881

Qy 1021 acctgattctgggttccaggaataacctggggagagttccagatccagcccgagtgct 1080

Db 1882 ACCTTGATTCGGGTTCCAGGAATACCTGGGAGATTTCAGATCAGCCCAAGATGTCT 1941

Qy 1081 gattcgaagtaagactcgcctgtcgaagtgaattactatgaaggcttttctatccc 1140

Db 1942 GATCTGCAGATAGACTCGCTTGTCAAGTGAATTAATGAGAGCTTTTCTATCCC 2001

Qy 1141 agttgactggtatctcaacctcgaagcgtgtgtattatgagagctcgaagtaagtg 1200

Db 2002 AGTTGTACTGATCTCACCTTCGAAAGCGTGTGATTTATGAGCTGCAGATTAAAGTG 2061

Qy 1201 cttaacaagaatggaactgtgtcctcaataattgattcctaagtggaacctatctg 1260

Db 2062 CTATTCAGAAATGACATTTGGTTTGCTAATTAATTGATTCTAAGTGAAGCCATCTTG 2121

Qy 1261 attaatgtcttggaagaagcttgaaagacagagagaagctgatacttggcctca 1320

Db 2122 ATTTAAATCTTGCGAGAAAGCTTGGAAGACAGAGAGCTGGATCTTGCCCTCA 2181

Qy 1321 gaaaagtattatgatgatagaatctcactaccacatca 1362

Db 2182 GAAAGTTATTATGATGATGAATCTCATCTACCATCATCA 2223

RESULT 3

AB037596 3508 bp mRNA ROD 20-OCT-2000

LOCUS Mus musculus Igt B mRNA for

DEFINITION beta-1,6-N-acetylglucosaminyltransferase B, complete cds.

ACCESSION AB037596

VERSION AB037596.1 GI:9650953

KEYWORDS beta-1,6-N-acetylglucosaminyltransferase B.

SOURCE Mus musculus (strain:ICR) 13 day embryo fetal brain

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)

AUTHORS Chen, G.Y., Kurosawa, N. and Muramatsu, T.

TITLE A novel variant form of murine beta-1, 6-N-acetylglucosaminyltransferase forming branches in poly-N-acetylactosamines

JOURNAL Glycobiology 10 (10), 1001-1011 (2000).

MEETING 20485361

REFERENCE 2 (bases 1 to 3508)

AUTHORS Kurosawa, N., Chin, G. and Muramatsu, T.

TITLE Direct Submission

JOURNAL Submitted (25-JAN-2000) to the DDBJ/EMBL/GenBank databases.

NOBUYUKI Kurosawa, Nagoya University School of Medicine, Department of Biochemistry, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan (E-mail: nkuro@tsuru.med.nagoya-u.ac.jp, Tel:81-52-744-2063, Fax:81-52-744-2065)

FEATURES

Source

Location/Qualifiers

1..3508

/organism="Mus musculus"

/strain="ICR"

/db_xref="taxon:10090"

/dev_stage="13 day embryo"

/tissue_lib="lambda ZAP"

/tissue_type="fetal brain"

1009..2214

/gene="Igt B"

/gene="Igt B"

CDS

1009..2214

/gene="Igt B"

/codon_start=1

/product="beta-1,6-N-acetylglucosaminyltransferase B"

/protein_id="BAB03495.1"

/db_xref="GI:9650953"

/translation="MGSWKYSLESLSLALMLMFWTDRKIMKHYHPPRAVNSIVLA
EVCLOMFGESEFYADSRKTTLENFTCPKIONNHTTETLSEEARPLATLTH
KYDPTFERLFRALIMPOVNYCVHVDSSATDFEKAVNQLLSCFPNRLASRMEPVYG
GESRLADLNCMDLVASKIPMKVLTNCGDPELTKNKEIVQYKRFIGNLTPVL
PPAAVGRITKYVHDELDRKRPVYHNTRAPKAPPHNLTIFYGTAVALRREFANFL
KDOSEVLISSKDPGSPDEFMYHTLNRIPEGVPMSPNMSMGNLAVRMDEAKH
GCGHRYHNGICITGNDLWILINSQSLFANKFELNTPYPLTVELERLNERLNQSE
IALDPSWF"

BASE COUNT 877 a 872 c 867 g 892 t

ORIGIN

Query Match 15.2% Score 207.2; DB 94; Length 3508;

Best Local Similarity 52.9%; Pred. No. 1.3e-39;

Matches 505; Conservative 0; Mismatches 428; Indels 21; Gaps 2;

Qy 337 taccagacctlaagaggtatgctcaaaagctgtctcaaaagagagaaagcttccca 396

Db 1231 TCAAGATTCAAAACACATACATACAGAGACTCTCGAAGAGAACCCGCTTCCCT 1290

Qy 397 atagcctattcttggtgtgtccaaagatgcaattatggtgaaagcttatcatgct 456

Db 1291 CTGGCTTACACTTACTACTCCAAAGATATGACACTTTGAGAGACTTTCAGGGCC 1350

Qy 457 atatacaaccagacacataattactgcatcacaatlgaatcgtaagcgacctataccctc 516

Db 1351 ATTTAATGCCCCAGAGACGTACTGTGTGACAGTGAGATAGTAAGCGCAGACACTTTC 1410

Qy 517 aaagtgtcgaatgaacaattagctgaagctctctcccaatttcatggttcccaatta 576

Db 1411 AAAGAAACGGGTGGCGGAGTTACTTAAGCTGTTCCCAAGTCTTGCCCTCTTGAGAT 1470

Qy 577 gaagcgtgtgaatgccccacattccagaccacagcgtgacttaattgctgttcggag 636

Db 1471 GACCCGGTGTCTATGTGTGGCTTCTCCGGCTCAGAGCTGATCTGAATGATGAAGAT 1530

Qy 637 ctctcgaagcttcaatccagatggaataatgtatcaacttctggtgggaagatttccc 696

Db 1531 CTGTGTGGCTCCAAAGATCCCTGGAAATATGTCTCAACCTCGGGCGAGGACTTCT 1590

Qy 697 ctgaagtcacaatttgaattggtgtcagaggttgaaanaactcaatggaagcaataatgt 756

Db 1591 CTGAATAACCAAGAAATAATAGTTCAGTACCTGAAGCGTTTATGGGAAGACCTCAC 1650

Qy 757 gagaaggtgaaacccccaaacagtaaatctggaagttacttaccatcatgaacttga 816

Db 1651 CCAAGGGTGTGTGCTCTCCACATGACATGGAAGAACCAAGTACCTCCACAGAGCTG 1710

Qy 817 cgggtgcttataatcatatggaagctcaataaagacaacatctccaagaagacacc 876

Db 1711 TTAGACCAATTAATAATCCCTACGTGC-----ACAATACAGCAAGATTAAGCTCCCA 1764

Qy 877 ccccaatacatcaagatatgttggcagtgcttatttggtttaagtaagcatgtgt 936

Db 1765 CTTACACACCTGACCATTTACTTGGACACTGCTTATGTGGCTCTCACAGGGAATTTGCT 1824

Qy 937 aaataatattcaacaactccatcgttcaagacttttggcctggtgctaaagacacac 996

Db 1825 AACTTTCTCTCAAAAGCCAGCGCTTCAGTAGAATTATCTCGCTGGCCAAAGACAGTAT 1884

Qy 997 tctctgataagcaacttgggtgctacacttgatccgggttccaggaataacctgggaagt 1056

Db 1885 AGTCTGATGAGCATCTTGGGTGACACATCAATAGAGATCCCTGAGACTCCCTGCTCAG 1944

Qy 1057 tccaagatcagcccaagatggtgtgctgactcgcagagtaagaactgaccttgcaagtgaat 1116

Db 1945 CCACCAACGCAATCTTGAGAGCGGTAACTCAGAGCTGTGAAGTGAATGACATGAAGCG 2004

```
OY 1117 tactatgaagctttctctatccagctgacatccaccttcgaagcgtatgtratt 1176
Db 2005 AAGCATGGAGGC-----TGCATGGTACTACACGTCATGCGATTGTGATC 2049
OY 1177 tatgaagctgcgaataaagctgacatacaagaatggacattggttgcataaatt 1236
Db 2050 TATGGAACGAGACTTCCACTGCTGATTAATTCGAAACCGCTTGTGCAACAATTT 2109
OY 1237 gattcctaagtgagacctatcttattcaaatgcttgcgacgaagaagcttaaga 1290
Db 2110 GAAGTCAACACATACCTCTTACCGTGAATGCTGGAATCTGAGACTTCGAGAA 2163

RESULT 4
AF318573 AF318573 108873 bp DNA VRL 14-FEB-2001
LOCUS Bovine herpesvirus 4 long unique region, complete sequence.
DEFINITION AF318573
ACCESSION AF318573
VERSION AF318573.1 GI:12802528
KEYWORDS
SOURCE Bovine herpesvirus 4.
ORGANISM Bovine herpesvirus 4.
Virus; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae.
REFERENCE 1 (bases 1 to 108873)
AUTHORS Zimmermann,W., Broll,H., Ehlers,B., Buhk,H.-J., Rosenthal,A. and
Goltz,M.
TITLE Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
JOURNAL Identification of an Origin of DNA Replication
PUBMED J. Virol. 75 (3), 1186-1194 (2001)
1152491
2 (bases 1 to 108873)
AUTHORS Zimmermann,W., Broll,H., Ehlers,B., Buhk,H.-J., Rosenthal,A. and
Goltz,M.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Project p24, Robert Koch-Institut, Nordufer
20, Berlin D-10555, Germany
FEATURES
source 1..108873
location/qualifiers
/organism="Bovine herpesvirus 4"
/db_xref="taxon:10385"
/misc_feature 1..108873
note="long unique region; LUR"
complement(445..633)
CDS
note="ORF B01"
/codon_start=1
/evidence="not_experimental"
/product="hypothetical protein pB01"
/protein_id="AAK07921.1"
/db_xref="GI:12802529"
/translation="MEGDVSVAFRRKRTRODLNNSAGNEVCAYLRKNHNSAGC
CAVKNRNSFHSLPGMPP"
614..4480
note="tegument protein; ORF3 BORF01"
/codon_start=1
/product="v-FGM-synthase"
/protein_id="AAK07922.1"
/db_xref="GI:12802530"
/translation="MDTSPSIOLCFFVSRAPLNSNECOVIOFTSDGOLPLTVLQY
BOYFVTVNGEBOAVRFVPPDLFSSGELTPRESSILKTDGRSLRFITGVSPA
PETALSRDLVALIQLSMGSLAPNSNLTQIKIYSLISIKCNLNPSPKIMSTLRAL
CPDPSLFRYKQSEIVHGAHFGHAPQGLSPFKGLDAFVGGFLNAGSGASGEMFT
POHQPOTIVNONGCTAVYIPVDYWIPELSTYLSGHMTDGLAVTGGOLVARNDC
LKFELYSNOMGNTAVYSSOLGIASTVSTLSSRSSELPYTRDINDAVYINNRCK
SLTCQGYEPIYESHLPDLSVEEGLIATGAVCESRNITDLPETRKINIVGLGT
FMSHDSGSEVNCILGALKYLSFLRPSOVGISAVKVCETMAKLLSVCGITGIGK
TYSALPPLNGLDLPETASNRQNMNTKOHNLIVATIVATAPDNTDVLK
AGCAIHGCOIKVLGRTOQPHITIVDKOEONVREKISIKRSPSPVLSITKILKO
ENMOSSMEMDILPSTYLLLELHFAVGCKDFVKAIDRLSSGVAAOQCGIGARDI
PIDVGLIVSDLSLAAPRGSENPWATAESLDLPLVNNEPVPGICSAIGCEPPLS
TYPPIGAHVAIVESLINTIVSAFRRKIDVTCTFNTWQTDLSHGLIELKRAEEL
CSOLGIGCVFTSCTSNRRGQKCLNLSLVRYLVITATAPCKDVTQGLTPDLKEPDSI
```

gene

CDS

CDS

CDS

CDS

```
VWLPISYHTFEGVISOLEPRDPSGCHINIDPLVYKILTYKHLINHESIISCHDV
GCGGLITACFEMAYAGGASIALTVPODDPVLFSTFPGLAWEPRVAKSTYOKHLE
TSDIYFVNGRTLPVSANFTVSRYKEHIFRESLINEVNWRFSTRKQRYPCET
OCTPEDPPKDLHLTFQPMCPHGPYKHOVANYLLPLGFTPPSLVLAEEAEFRV
LSTPTDKVATVITDITNVEFGICLANIDDALGDAIKAMTKHNSVLADEKTLI
NSPDYFSLAIGHTACOLLENNKFMGYNKPSNTMKCKENSGIMESRMILNFIENH
AVALDSMGSLPLPGIOCTHGLAFAPSTYEMLSHGMVATOFYAGDIAISGALTYP
ONPTGYVITSGICSDGRHLALHDPLGSLNNLMQPHILPKATPPLAKSPKMRMLDLH
IWANKYREMDPPPHDPDLNRIKYM"
4608..4811
note="ORF B02; BORF02"
/codon_start=1
/evidence="not_experimental"
/product="hypothetical protein pB02"
/protein_id="AAK07923.1"
/db_xref="GI:12802531"
/translation="MASPTQYCHRMVRYHGPRKSAKISYGYKRDYTHQLSKNSAYS
GMSPOPCSCADLNKLFVSQCPIS"
5286..5513
note="ORF B03"
/codon_start=1
/evidence="not_experimental"
/product="hypothetical protein pB03"
/protein_id="AAK07924.1"
/db_xref="GI:12802532"
/translation="MGFNPMVLYFRPCNCTPTSLKHGLYIKYIKSYLIRATFVRYV
HIGPPKRDTRPPGPGPPLSLVORLAKK"
5943..9347
note="ORF 6"
/codon_start=1
/product="single-stranded DNA-binding protein MDPB"
/protein_id="AAK07925.1"
/db_xref="GI:12802533"
/translation="MAKTAAGGOYEDNIGSKAPICPGCIYVYKKEPPTKEASLG
NGPPGVSWSVPLRLGLTVPPGPRVAKVHKIDDTYSAVTSYHREVMESNDC
FTFTHGTGIPOLCOETRDLEGTTHSPNSNTNLSIDLCGTGDCIDEMILGYVIT
EFGFERLYHSHMPLIGHVEVCIHQHAKRKPLVDEDLFSLIARDMKRHSVDS
QYLENSLYTAIAQAIKVDVATVQMSQFVNDQMKRKYVAKKDPRESSSRGPGGL
SLIMDSVASELAVSYGLFIDAPQMSALDTSWPIVDCTEDRQLQALANAKL
OSIHNTOLESTSMYLTIRIQONKPKTKGDLNVSYYLQHGSLYSEADQDEGQ
PVFGOVSNLSGSTYTIHLAAASMCPTLARKCYVLFQCHQKSTQNSINISNY
VGRASDMDNCLGGKRPACINTFLPRDRPPTITTSRRDPYITGSGWYNDL
ILGFRASPREKDEBQGVDEQVCKYTWOLITLLELKGIDTTPREKLVADISF
IKVREDIDAVDIEVLFINKCKNNINRETIKSHHYVAKVONCFAPPCVGLT
YKXSLTIIDVCLPCMTEQDNSSMGWPLMHNITLTNRKGCCTPBGVYTG
AELKVMHSDPCDFEVDAINIGISPTKQVRISSAMVPEVTIKIRNIFSSNSG
SEAIOSGLKSGSKKNYIVTGYMKFLNHYRIMPEOTKMSLFMWHFSTKQIPL
ITGVKDLVALANYLEYNSKLSIEIDLDIIDNLTAKILNNAIILACQTOFY
STPHCLVPTVRYVPGEEPHVLENEEFLGTEQYLVKQNRQAIVQALIKEDVAQMG
KLRIITVPMVINKYNGINGNNGIFGNGIGYMGAGVDENLIFEANAPKROSTNAM
KRKHVMTPIYDNLIKRTYTPSTSPVENIRRIDNLEDKRONNIFRDVYELVKG
LGQACADLTADDLOFYLGERTYINSDLSLOSISDAGVWCQSDVSPFLGFTVSEE
QLEPIGLEQDTSPASIDEDFFQASGLSTIAGKRRKLNCLSPDL"
9352..11430
note="ORF 7"
/codon_start=1
/product="transport protein"
/protein_id="AAK07926.1"
/db_xref="GI:12802534"
/translation="WAEFLAIVSYQVDFLSKTSLSLTFCDPRHIDGAPORNKRLTL
LIDLAYSHLVQONTSCSLSELEHLINITYITWQLNVECLNNTYERKPYFMHLHS
DLCCFHTVINCLEFGDCVPIPSLIIINDIEFBSRLVSEYCSRNAADLGEETIN
LGTLRNTPSLPLDLPYPSISCMKCNQLESTDRIPNAGNFTIINMSICNTAVAPAE
PVQGLFENEELHNLINPDGCTNKQLESTDRIPNAGNFTIINMSICNTAVAPAE
NLITYNSGGSIBKISCSQMATLTIHERMASRQKVASHSALTTPYHPFDYSYA
YPSLSEFGGIGFYIIMNIOALCKDQSCSTPLAKSNYKTEIKERKEIVLRNLLAPB
TREBIDDPNPSISQARHTIGETSEQISDAAKRNNYLQKCTEGCLKIDDKDET
QKILMNCNLTLRITGMVYKAAALIKHFMFKRFFILPQHNHYTINELYESKRI
KNSLYVQLSQEHDSITQLYSLILGPEGPPFSPPIILLANCLDAKVAHPHK
MLISEVIMPSMEKMDIADENAFYNTIGTKLNSVOOCAMFYIELVMSVLNRWE
KQINIVSSTLSLPHNLFSKPTFRSGIYIYEKTAAPLFIHNTGWIIFKDYALLVHH
LQLSHKRYVL"
11420..11404
```

CDS	<p> <code>/gene="gb"</code> <code>11420..114044</code> <code>/gene="gb"</code> <code>/note="ORF 8"</code> <code>/codon_start=1</code> <code>/product="glycoprotein B"</code> <code>/protein_id="AAK07927.1"</code> <code>/db_xref="GI:12802535"</code> <code>/translation="MYKTLFPALIKVCSFNQTTHTSTTSPISSTSTSTSTSK PNTSTINSLSASPONTSTKSPSTDNQSTPTLPVTADSNFKRYRCSASS SGELFREDQCPDPTDKKHVEGILLVKNIVYIFKRYRKIAVSIVYRMSO AAVNRDDISRAIPYNEISMDIRYCFESAMATVINGIINIDDSNRKSVLPVA GLTENINRYSOPLIYAEPGPGIYRVATVTCVDEYVAVSAPYHFEALDTI ELSPRCHNSOCTHGTQSRDAKVIENHQTVDYERGHPTKXRIEIKDEETIS WREDERAICDPVTKTTPRAIQTTHNSFPFVANEVYASLTINQSEETRGNETI LNCNSTINETEYVTKFNKSHIDGKVKYIKINGLEFLINQAKPLNLSHTNVTI ERNKTKGNSRKRQSVDTKFOGAKGLTAQVQYAVDHLRTSMNHILELITWCREQ KKNLMWELSKINPVSAAYGKPVAKAGDAEMVSECINQVQASVNIKSRMD DPKCYSRPLVTFKFNSTATFRGOLGTENEILLNTHECTRPTADHYFEVKNATHY FKRYFKVKTMDNNISTIDPLNLNTEIDNTDEFTVYSETERRMSADLETFR EYKYVQKATREDIDNTIDLNQDQVYDLSMMADGIDGKVVNTNFGSVTVYGS IVGQVSPFTNPIGVTTILLIVYFVPIYSRRNNKNEAPIKIYIPNTRKASEQ NIQPLGEETKRILLGMDHQDQSEHGKSEESHKFGLEQLLDGLDQLLRNGYTRLP TFDPSGNDSETHOKY"</code> <code>11484..17201</code> <code>/gene="DPO"</code> <code>14184..17201</code> <code>/gene="DPO"</code> <code>/note="ORF 9"</code> <code>/codon_start=1</code> <code>/product="DNA polymerase"</code> <code>/protein_id="AAK07928.1"</code> <code>/db_xref="GI:12802536"</code> <code>/translation="MEFNPIYSKKTKDKRISCVDDPKKTNPCYHLCVKECTPTGPT GVSITTDTPVLFHODKEPIYFGSDHORSYSAKPPAPIPKDKLFPHYDIE TIVPRPCPDPIPOFOTDIPNGTVLKLKGDQDASVCVNVKTSAPPEINQVNLN LSYVTOOTLNGVKNQCTKSTTQEKRIKTEGPELAPYVKTLSAVEINQVNLN TSCGCEVESNVASTFTIDNKFSTFGWYSCNPPRISQSRSDPDLDEDCGLDQ FHEEDQMPYTIIMSPDIEICIGQGFCAKDELDVQISCIIVIVCSBPNNILS LGTCDPYANDVYFESLDMFAFTLLDEFDIVGTNAINFPPYIIDATVY NIPRHTKVISGSTEVENQPMNTGAFNRSKIRISGVPIDMVOYQDKLSIDY KLNTVAHCKGTQSEDEVYKEIPLFRSGAGARIGSYCVLDSVVLDLKFMTHV EISLIAKMAKIPARVILDGOIIVFSCLEAKKEHFIIPGAQPGAGCATVIN PIREYTPPLIVDFALYPSIIQAHNLCSVMQIDONLHLHKDDDETHTLSTGP IHEVQKRTSLSLTLTAWAKRKTIRKLANCDDGPKTIIDKQOLAKYCNVY GFTGVASGILPCPIAETVTLQGTMLEKSAVEKITPRLSDIVSIFVPCDPMSE RVIGVDSDSEIECRGPMQSVNFADKLETETKALEKPILEAKETFOCLMLTK KRYIGLISDKLVKGVDLIRKACSVQTTSEILDLVREDEVQAOAYLROQPA KVYSDGLPVGFWKVIDVINSYALATGKVSSEILSFSTELSPFGGKTTLPHLTV YRKISNRELPOIHRIPIYVFIKDGQRGCKDAEDPTVSONKPIPISEIYFDKLI HGVANITQCLFGNSNMTVELINFPVNIPIPSF"</code> <code>17287..18577</code> <code>/note="ORF 10: BORF1; conserved in other gamma-herpesviruses"</code> <code>/codon_start=1</code> <code>/product="unknown"</code> <code>/protein_id="AAK07929.1"</code> <code>/db_xref="GI:12802537"</code> <code>/translation="MSTSLVFESEIGWAVCNCTIITMTNKIDIPVGGVGMVKKV FLPLSLNOLINNYISFGLVOSIFELEDAVOCNTALVOCRRPIRTPLEPIHYVEN MLPLFVYTONNTIINSLSLAILFMRPTCTGICSIISAPPTIIPQSLAHLES TIDQGLILSGAICKNKTYMTFRGPPSOFMEHTMNSGRDLDVADCAVNSVD IKINIMEVNPNGLYKVSNTVNPBKPDHILWIKVAKPDLVAVSLPSTLTS CGKWNITDESEKLIKPELTNLKATYTGNSSTLAIVITGTNNPWTLEPTIM LPMPIQVTKNPTNITTIKKDLAIAQVYGTSTLDRQPPASVYFNPQDILITW EDSMNVATIGENIYSRCHLNKSENPSPMDP"</code> <code>complement(18948..19345)</code> <code>/note="ORF Bof; short ORF of immediate early transcript 1 RNA"</code> <code>/codon_start=1</code> </p>	
Query Match	14.3%; Score 194.8; DB 58; Length 106873;	
Best Local Similarity	52.7%; Pred. No. 1.6e-36;	
Matches	520; Conservative	0; Mismatches 427; Indels 39; Gaps 3;
Qy	316	atgaccagtgatctgtgacattatcagactctaaaggtatgctcaaaagctgtctca
Dy	107637	ATMAACGAGAGCTGTGACGCGCTTCAAGGCCCAAGAGATTGATCATGTTCACAGT
Qy	376	aagagagaaagactctccaatagccattcttctgtgtcccaagaagatgcaattag
Dy	107697	AAAGAAGAGCTAGACTTCCCATTCCTACTCGATGCGGTGAGAGATGGAATGGAAC
Qy	436	gttgaaggtatccatgctatatacaaacagaacaatatctcgtcatccattatgat
Dy	107757	TTTGAACGGCTGCTCGAGCGGTGTATCCCTCAGAACATATCTGTGTCCAGGTGAT
Qy	496	cgtaaggaacctgataccttcaaatgtccatggaacaattagctaaagtgtctccaat
Dy	107817	GTGAAGTCCCGACAGACTTTTCAAGAGCGGTCAGAGCCATTATTCTGCTCCCAAT
Qy	556	atttcattgcttccaaattagaagctgtggaatatgccacattccagactccaggtc
Dy	107877	GCTTCATGCGCAAGTGTGTCCGGTGTATTCCTGCTGCTCCAGAGTCAAGT
Qy	616	gatttaattgtctgtcgagactctcgagctctgaagctcgaatccagtgaaatgatac
Dy	107937	GACCTGAAGTGTATGAGACACTTCTCCAGAGCTCAGTGCATGGAAGTACTTACTGAT
Qy	676	tgtgtgagcaagatttcccttgaaagtaatttgaaattgtgtgtcgaggttgaaaaa
Dy	107997	ACATGCGGACAGACTTCCCATTAAGACCAATGCGGATGTCCTGCTCCAGATG
Qy	736	ctcaatgagcaaatatgtgtgagagcgtgtaaaccccaagaatgaaattgaaagattc
Dy	108057	TTGAAGGTAAGACGTAATGAGATCTGAGTACCTTCTGAGTCCAAAAAATGCTGCG
Qy	796	acttaccatcagaacttagacggtgtccttataatgtaagctcaacaataagaca
Dy	108117	AAATACCGGTATGAGGTGACAGACACTGTACCTT-----ACG 108155
Qy	856	aacatctcaagaagacaccccccaataactgataattgttgagcagtgctatttt
Dy	108156	AGCAAGATGAGAGACCTCCCTGTAATTTACCATTTTCACAGGAGAAATCCATTTT
Qy	916	gttttaagtaagcatttgttaatatatttcaacaactcgaatcgltcaagaacttttt
Dy	108216	GTGGCTCTGAGACCTTTGCCAATGCTTACAGAACCTTAATCCCAAGACTGTGT
Qy	976	gcttgatctaaagaacatactctctgtatgacacttttgggttaactgtatcggttt
Dy	108276	GAAATGGTCAAAAGACACTATAGCCCGGAGAACACCTCTGGGACACCTTACAGCTGCT
Qy	1036	ccaagaaatacctggagaaatttccagatcagcccaagatg---tgtctatctcagaagt
Dy	108336	CCGTGATGCTGTGCTGTCTGTCTAGCACACCCCAAGTATCACATGACATGACTGCC
Qy	1093	aagaatcgactgtcgaagtgaattactatgaagctttttc-----tat 1137
Dy	108396	ATGGCCAGGCTGCTCAAGTGTGACGACAGAGGAGATGTCCATGCGAGGCGGCTTAT
Qy	1138	cccaagtgactcagatctcaccctcgaaagcgttgtcatattatgagctcagaataaag
Dy	108456	GCACCTGCTCTGGAATCCATCGGAGGCGCATGTGCTTTAGCGGGCCGGGACCTGTAC
Qy	1198	tgccttatacaagaatgaaattgtctgataataaattgattctaaagtgtgacactatc
Dy	108516	TGGATTCTCCAGAACCTACCTCTTGGCATAAGATTGACCGGAGGTGATGATTAAC
Qy	1258	ttagtaaatgctgtgagcagaagaagct 1283
Dy	108576	GTCTGCGAGTCTTAGAAGAGTACT 108601
RESULT	5	
AB037597		

LOCUS AB037597 1206 bp mRNA ROD 20-OCT-2000
 DEFINITION beta-musculus 180A mRNA for
 beta-1,6-N-acetylglucosaminyltransferase A, complete cds.
 ACCESSION AB037597
 VERSION AB037597.1 GI:9650955
 KEYWORDS beta-1,6-N-acetylglucosaminyltransferase A,
 source Mus musculus cell_line:F9 cell CDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)
 Chen, G.Y., Kurosawa, N. and Muramatsu, T.
 A novel variant form of murine beta-1,
 6-N-acetylglucosaminyltransferase forming branches in
 poly-N-acetylglucosamines
 Glycobiology 10 (10), 1001-1011 (2000)
 JOURNAL 20485361
 MEDLINE 2 (bases 1 to 1206)
 REFERENCE Kurosawa, N., Chin, G. and Muramatsu, T.
 Direct Submission
 Submitted (25-JAN-2000) to the DDBJ/EMBL/Genbank databases.
 AUTHOR(S) Nobuyuki Kurosawa, Nagoya University School of Medicine, Department
 of Biochemistry, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550,
 Japan (E-mail: nkuros@suru.med.nagoya-u.ac.jp, Tel:81-52-744-2063,
 Fax:81-52-744-2065)
 FEATURES
 source
 1..1206
 /location="Qualifiers"
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="F9 cell"
 1..1206
 /gene="Tgnr A"
 1..1206
 /gene="Tgnr A"
 /codon_start=1
 /product="beta-1,6-N-acetylglucosaminyltransferase A"
 /protein_id="BAB03496.1"
 /db_xref="GI:9650955"
 /translation="MPSVRYVFIVSVTVIVPILVLSFGDSYOKLINSVRL
 SOVCSFIDGSRFLRMKIMHERKPSCTEVVTSHTPLASLEVDPLAIVYVH
 HNFDIFARLFAIPONTICVHNDEKTAEFKAVEDLYSCFNALASMPVYVGS
 GISRLQADLNCIKIDLTSEVPWKYAINCGDPLKTKKEIVYLKIGKRNLPGL
 PPAHAGRTYRVHREHLSKELSYIIRITALKPPPHNLTIVGSAVYALSEFNELV
 RDPRAVDLHMSKDTFSDEHFVWTLNRIQVPMSPMNAWGTNLRAVKMDWEAH
 GGCHGVHGHGICIVGNDLQWLINSOSLFANKFELNTPPLVECLERLRERTINOS
 IAIOPSWYF"
 BASE COUNT 300 a 318 c 285 g 303 t
 ORIGIN
 Query Match 14.3%; Score 194.4; DB 94; Length 1206;
 Best Local Similarity 52.1%; Pred. No. 1.5e-36;
 Matches 497; Conservative 0; Mismatches 436; Indels 21; Gaps 2;

QY 637 ctctgaagctctcaatccagatggaaatattgtatcaactctgtggcgaagatttccc 696
 DB 523 CTGTCCACCTCCGAGAGTCCCTGGAGTAGTACGCCATCMAACCTGTGGAAACACTTCC 582
 QY 697 ctgaatcaaatatttaatttgctcagagctgaaanaactcaatggagaatattgtg 756
 DB 583 CTTAAACACACAGAGAGATAGTTCAGTACCTGAAAGGCTTAAGGGAGAACCTCACT 642
 QY 757 gaacgctgaacaccccaacagtaaatlgyaaagattcaactaccatgaactaga 816
 DB 643 CCGGGGGTGTGCTCTCAGCCACGCCATTGGAAGGACAGGTAC-----GTCCACCGG 696
 QY 817 cgggtgccttaataatattgaaagtcaccataaagagaacacatcccaagaagacc 876
 DB 697 GAACACCTAAGCAAAAGAGCTTCTCAGTATGAGAACACAGGCTTGAAGCCCCACT 756
 QY 877 cccataacatcaagatattgttgcaagtgctatttggitttaagtaagcattgtt 936
 DB 757 CCCACACACTACCAATTATTGCTGCTATGTCGCCCTGCGAGAGATTGCT 816
 QY 937 aataatattcaacaactcaatcagttcaagacttttgcctggtctaaagacataac 996
 DB 817 AACTTGTCTCCGTGACCCACGGGGGTTGATTGCTCCATTGTCCAAAGACTTTC 876
 QY 997 tctcctatgaacacttttggtcactatgcctgaggttccagaaataacctgggaagt 1056
 DB 877 AGTCCGATGAGCATTTCTGGTGAGCCTCAATAGATTCCAGAGATCCCTGCTCCAG 936
 QY 1057 tccagatcagcccaagatgtgtctgactcgaagtaagctgccttgcagtggaat 1116
 DB 937 CCACCAACCGCATCTTGAGCGGTTACCTGAGCTGTAAGTAGAGATGGACATGGAAGC 996
 QY 1117 tactatgaagcttttctatccagttgactgactgactacactcgaagcgtgtgatt 1176
 DB 997 AAGCATGAGAGC-----TGCCATGCTGACATACGTCATGCAATTTGATC 1041
 QY 1177 tatgaactgagaatgaagtggtgctatacaagaatgagacattggttgcataaatt 1236
 DB 1042 TATGAAACGAGAGCTTGAGTGCTGATTAATTCGCAAAAGCTGTTGCTAACAAATT 1101
 QY 1237 gatcctaagtgagaccatctcttgaatgaatgcttgcagaaagcttgaagaa 1290
 DB 1102 GAATCAACACATACCTCTTACCGTGATGCTGGAAGCTGAGGCTTCAGAA 1155
 RESULT 6
 AX087948 1317 bp DNA PAT 17-MAR-2001
 LOCUS AX087948
 DEFINITION Sequence 14 from Patent WO0114535.
 ACCESSION AX087948
 VERSION AX087948.1 GI:13396926
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1317)
 Schwiientek, T. and Clausen, H.
 UDP-n-acetylglucosamine: galactose -4(1,3)-n
 -acetylglucosamine -9(a)-r/(glucose to galactose) -9(1,6)-n
 -acetylglucosaminyltransferase e, c2gnf3
 Patent: WO 0114535-A 14 01-MAR-2001;
 Schwiientek, Tilo (DK); Clausen, Henrik (DK)
 JOURNAL
 FEATURES
 source
 1..1317
 /location="Qualifiers"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..1317
 /note="The gene of C2GNF2"
 /codon_start=1
 /protein_id="CAC34676.1"
 /db_xref="GI:13396927"
 CDS
 QY 577 gaagctgtgaataatgcccaattccagaactccagcgtatattgaattgcttgcgac 636
 DB 463 GAGCCGGTGTCTATGTGAGATCTCCGGCTCCAGGCTGACCTGAATGATCAAGAC 522

```

BASE COUNT      359 a      307 c      322 g      329 t
ORIGIN
/translation="MVMOKRLCOLHYLMALCGLTALVAKLISFRLKLCOSDHLGLES
RESOSCRNLINFLNFKLPAKRSINGSCVGGDEAVLQALINLVPKKRKEPPTDTH
YLSITRCEHNEFKARKKTIQPLSCKEEVPFLAYSMVIVYKENEERILLRAYAPONIY
CVYHDEKSPETFEKVAIKATISCPFNVPFLASFLVYVYASMRVADLNCMEDLSSV
PWKFLNTGDTDFIKSAENVOALMKLNGRNSMSEVPPKFKETRMKTHIEVYRDTL
HLTKRKDDPPNYLMTFETGNAYIVASREIYVWQHKEPQSDQLIEMVADTSPSDHLWAK
TLQARWMMGCVSPNNPKYDIDMTSIRFAYVWQHGEIDGAYAPDSGTHORACVHLN
YAGADLNMMLQNHHLANKFPDKVDNDALQCLEEYLRKATAYGTEL"

```

Query Match	Similarity	14.1%	Score 191.8	DB 10	Length 1317
Best Local	Similarity	52.5%	Pred. N0.64e-36		
Matches	517	Conservative	0	Mismatches 433	Indels 36
					Gaps 3
Qy	317	tgacagatgattgacattatcagactctaaaggtatgctctaaagctgtctcaa	376		
Db	317	TCACGAGAACACTGAGACCTTCAAGGGCTGAAGAGATTCTATACAGTTCACACAGCA	376		
Qy	377	aggagagaaagctctcccatagctatctcttggtgtgcacaaagatgcaattatg	436		
Db	377	AAGAAGAGGTGAGTTCCTATTGCTACTCTATGCTGATTATGGAAGATTGAAACT	436		
Qy	437	ttgaaaggtctatccatgctatatacaaacagacacataattactgcatctatgac	496		
Db	437	TGGAAGGCTACTGCGAGCTGTGTATGCCCTCAGAACATATACGTGCTCATGTGATG	496		
Qy	497	gtaaggcactgtatacctctcaaaagtgcacatgaacatttagctatgctcttccaa	556		
Db	497	AGAAGGCCCAAACTTTTCAAGAGCGGCTCAACAACTATTCTTCTCTTCCCAATG	556		
Qy	557	tttcatgtcttccaaattagaggctgtgtgaatgtccacatcttccagactccagctg	616		
Db	557	TCTCTATACCCAGTATGCTGTGGGTGGTTATGCTCTCTGCTCAGGTTGCAAGCTG	616		
Qy	617	atttaattgtctgtcggacctcgaagttctcaatccagtggaatatgttatcaact	676		
Db	617	ACCTCAACGCAATGAGAAAGCTTGCTCAAGCTCACTGCGGTGGAATCTCTGTAATA	676		
Qy	677	tgtgtgtgaagaatttcccccgaagctcaaatattgaattgtgtcagaagtgaaaaac	736		
Db	677	CAGTGTGGAGCACTTCTCTATTAAGACATCCACAGATGTCACAGCGCTCTAACATGT	736		
Qy	737	tcaatgagacaatatgttggagagcgttgaaccccccaacagtaaatgtgaagaatca	796		
Db	737	TGCAATGGAGGAAATACATGAGAGTCCAGAGTACTCTCTAAGCAAGAAACCCCTGTGA	796		
Qy	797	cttacaatcatgacttagacgggtgctctatgaatatgtgaagctacccaataagaca	856		
Db	797	AATATCACTTGTGAGTAGAGAGACACATTAC-----CTAACCA	858		
Qy	857	acactccaaaggaagacccccccaatacaatcaagatattgtttggcagtgcttattg	916		
Db	839	ACAAAGAGAAAGATCTCCCTCTTAATATTAACTATGTTACAGGAAATGCTTACATG	898		
Qy	917	ttttaagctcaagatttgttaatatattttgaacactccatcgttcaagactttttg	976		
Db	899	TGGCTTCCCGAGATTGTGTCCAAACATGTTTGAAGACCTTAATCCCAACACTGATTG	958		
Qy	977	ccgtgactaaagacacatactctccatgtagacatttggcttactctgtatcgggttc	1036		
Db	959	AATGGGTAAGACACTTATATAGCCCATGATACACCTCTGGCGACCTTCACGCTGCAC	1018		
Qy	1037	caaggaataactgtgggagatttccagatcagcccaagatg----gtctgacttcgagat	1093		
Db	1019	GGTGATGCTGTGCTTGTCCCAACACCCCAAGTGCACATCTCAGACATGACTTTCTA	1078		
Qy	1094	agactcgcctgtgcaagtgaattactatgaagcttttctatcccaagt-----	1143		
Db	1079	TTCGCAAGCGTGTCAATGTGCGAGGGTATAGAGGAGACATCGAATAAGGTCCTCTATG	1138		
Qy	1144	-----tgtactgatactcaactccgaagcgtgtgtatttgaagctgcgaattaaagt	1198		

QY	1139	CGTCCCTGCTCGTGGATATCCAGCGGGCTACTCTCGTTTAAAGGGCTGGGGACTTGATT	1139
QY	1139	ggctatcaagaagcggacattggtctgctaataattgattctcaagctggaacctatct	1258
Db	1199	GGATGCTTCAAAACCATCATCCTGTTGGCCACAAAGTTTGACCAAGGTAGATGATATG	1258
QY	1259	tgattaatgcttgccagaanaagct	1283
Db	1259	CTCTTCAGTGCCTTAGAAGATACCT	1283
RESULT	7		
LOCUS	AF038650	1317 bp	mRNA
DEFINITION	Homo sapiens core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase (C2/4GNT) mRNA, complete cds.		
ACCESSION	AF038650		
VERSION	AF038650.1	GI:4511880	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 1317)		
TITLE	Schwientek,T., Nomoto,M., Levery,S.B., Merx,G., van Kessel,A.G., Bennett,E.P., Hollingsworth,M.A. and Clausen,H.		
JOURNAL	Control of O-glycan branch formation. Molecular cloning of human CDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase		
FEATURES	1. Biol. Chem. 274 (8), 4504-4512 (1999)		
source	99143102		
gene	2 (bases 1 to 1317)		
CDS	Schwientek,T. and Clausen,H.		
	Direct Submission		
	Submitted (28-NOV-1997) Dental School, Faculty of Health Sciences, Copenhagen University, Noerre Alle 20, Copenhagen 2200, Denmark		
	Location/Qualifiers		
	1..1317		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	1..1317		
	/gene="C2/4GNT"		
	1..1317		
	/gene="C2/4GNT"		
	/note="core 2/4-GNT"		
	/codon_start=1		
	/product="core 2/core 4		
	beta1,6-N-acetylglucosaminyltransferase"		
	/protein_id="A021525.1"		
	/db_xref="GI:4511881"		
	/translation="MVMKRLCQCHYIWLALCYMLATVALKLSFRKCDSDHIGLES		
	RESOSQDEHNLNFKLPARKSINCSVTGSDGEAVIALINLELVKKRKEPTDTH		
	YLSITRDECHNLKAEKRFQPLSKSEVEFPYASVNIHEKLENERLRAVYAONTYH		
	CVHVDKSEPEFEKEAVKALISCPFVFIASLIRVYVSMGRVADIMCDMLDLOSSY		
	PKWELTNGTGDPIFKSAENVOLKMGNSNSSEVYPKHEKRMKYEVRDITL		
	HLTPKRRPPYNIITMGKDIADISDMTSIALVYASRFGVHVLKNPSQDLEWVXDYSPEDELMW		
	TLQRRKMPGSPVNPRIKGDIDISDMTSIALVYASRFGVHVLKNPSQDLEWVXDYSPEDELMW		
	YGAQDLNMLNHLANKFDPKDDNADLOCLLEELRKALITGTEL"		
BASE COUNT	359 a	307 c	322 g
ORIGIN			329 t
Query Match	14.1%	Score 191.8	DB 88; Length 1317;
Best Local Similarity	52.5%	Pred. No. 6	4e-36;
Matches 517; Conservative	0;	Mismatches 433;	Indels 36; Gaps 3;
QY	317	tgacagtgatgtgacatttcaagactctaaagaggttaagctcaaaaagcttgctctca	376
Db	317	TCACCGAGACGCTGTGCACTTCACGCGAAGCAAGTTCATACAGTCCACCTGAGCA	376
QY	377	aggagagagaagctcccaaatgactatcttggctgtgccaaaagatgcaattatg	436
Db	377	aggagagagaagctcccaaatgactatcttggctgtgccaaaagatgcaattatg	436

[illegible]

Db 377 AAGAAGAGGTGGAGTTCCTTATTCATACATCTATGTTGATTCATGAGAAGATTGAAACT 436
 Qy 437 ttgaagagcttcatctatatacaacccagacacatattcatgcatcattgac 496
 Db 437 TTGAAGAGCTACTGCGAGCTGTGTATGCCCTCCACACATATACGTGTCCATGTGAGTG 496
 Qy 497 gtaaggaacattgataccttcaaaagtgtccatgaagacaatttagtaagtgtcttcaata 556
 Db 497 AGAAGAGCTCCAGAACTTTCATCAAGAGCGGTCACAAAGCAATTATTTCTGTGCTCCCAATG 556
 Qy 557 ttctatgtctccaaattagaagcgtgtgaaatgatccacattccagactccagcgtgt 616
 Db 557 TCTTCTACCCAGTAACTGCTGTTCCGGTGGTGTATGCTCCCTGCTGCTCCAGGAGTGAAGTG 616
 Qy 617 attaatgtcttgcgagccttctgaagcttccatccatccagtggaataatgttataact 676
 Db 617 ACTCAACTGATGGAAGAGTCTGCTCCAGAGCTCCAGTCCGTGGAATTAATCTCTGAATA 676
 Qy 677 tctgtggaagatttcccccgaagtcacatttgaattgtgtgcagagtgtaaaaaac 736
 Db 677 CATGTGGAGGAGCTTCTTATTAAGAGCAATGCAAGATGCTCCAGGCTCTCAAGATGT 736
 Qy 737 tcaatgtgaacaaatatgttggagacggtgaaaccccccaacagtaaatvgaaagattca 796
 Db 737 TGAATGGAGGAATATGATGAGTCAAGAGTACCTCTTAAGCAAAAGAACCCGCTGGA 796
 Qy 797 cttaacatcatgacttaagaaggtgtccttaagaattgtgaagctaccataaagaaca 856
 Db 797 AATATCACTTGGAGTATGAGAGACACATTAAC-----CTTAACA 838
 Qy 857 acatcccaagagagacaccccccaacacacacacacacacacacacacacacacacacac 916
 Db 839 ACAAGAGAGAGTCTCTCCCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 898
 Qy 917 tttaagtcacagcatttgttaataatatttcaacaacacacacacacacacacacacacac 976
 Db 899 TGCTCTCCAGATTTGCTGTCACACATGTTTGAAGAACCTTAATTAATTAATTAATTAATTA 958
 Qy 977 cctgtgttaagagac 1036
 Db 959 AATGGGTAAGAGACACTTATAGCCACAGATTAACACCTCTGGGACACCTCTCAAGCGTGAC 1018
 Qy 1037 caaggaatcctgggagattccagatccagccagatg----tgtgtatctgcagagta 1093
 Db 1019 GGTGATGCTGCTGTGTCAC 1078
 Qy 1094 agactgccttgcagagtgatcttaactgaagccttttccatccagf----- 1143
 Db 1079 TTGCGAGGCTGTGATGATGAGGATGATGAGGATGATGAGGATGATGAGGATGATGAGGAT 1138
 Qy 1144 ----tgtactgtatctcacccttcgaagcgtgtgtatttatgagctgcagaattagtt 1198
 Db 1139 CTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1198
 Qy 1199 ggccttcaagaatgagacattgtgttctataataattgtatcttaaggtgtgacccatct 1258
 Db 1199 GGATGCTTCAAAACCATCTGTGTGCGCAACAAGTTTGAACCAAGGTGATGATGATGATG 1258
 Qy 1259 tgattaaatgtgtgcagaaagct 1283
 Db 1259 CTCTTCAGTGTGATGAGAAATACCT 1283

RESULT 8
 AF102542 2128 bp mRNA PRI 31-JAN-1999
 LOCUS Homo sapiens beta-1,6-N-acetylglucosaminyltransferase mRNA,
 DEFINITION complete cds.
 ACCESSION AF102542
 VERSION AF102542.1 GI:4204683
 KEYWORDS human,
 SOURCE Homo sapiens
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 CDS
 BASE COUNT
 ORIGIN
 Query Match 14.1%; Score 191.8; DB 88; Length 2128;
 Best Local Similarity 52.5%; Pred. No. 6.6e-36;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

1 (bases 1 to 2128)
 Yeh,J.C., Ong,E. and Fukuda,M.
 Molecular cloning and expression of a novel beta-1,
 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I
 branches
 J Biol. Chem. 274 (5), 3215-3221 (1999)
 99115671
 2 (bases 1 to 2128)
 Yeh,J.C., Ong,E. and Fukuda,M.
 Direct Submission
 Submitted (28-OCT-1998) Glycobiology Program, The Burnham
 Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037, USA
 Location/Qualifiers
 1..2128
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="brain"
 /dev_stage="fetus"
 354..1670
 /note="glycosyltransferase: core 2 Gnt-M"
 /codon_start=1
 /product="beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="A010824.1"
 /db_xref="GI:4204684"
 /translation="MVQMRKLCQHLVLMALGCMYMLATVALKLSPRLKDSBHLGLES
 RESGQYCNHILYFLKLPARKSINGSGVAGQAEVLAHLNLEEVKRRREFDTTH
 YLSITRDCENHILYFLKLPARKSINGSGVAGQAEVLAHLNLEEVKRRREFDTTH
 CVHYDESPETFEKAYKALISCPNPFIASKIYRVYASMPQADLNMDLQSSV
 PWYFLMTCTDPIPKSNAMVYLAASDQVHLKPKSQLLLEWKDYSPDEHLM
 LIOARMPGSPVNHPIYDISDMSISIRLVKMGHGDIDGAPVPCSGSIHQRAICV
 YGADLMMWLNHILANKPDPYDVALQCLEEYLRKAIYGTED"

317 aggaagagaagagcttcccaatagcctatcttgggtgtccaaagaatgcaatg 436
 Db 670 TCACCGAGACTGTGAGCACTTCAAGCTGAAAGAAATTCATACAGTCCCATGAGCA 729
 Qy 730 AAGAAGAGGTGGAGTTCCTTATTCATACATCTATGTTGATTCATGAGAAGATTGAAACT 789
 Db 437 ttgaagagcttcatctatatacaacccagacacatattcatgcatcattgac 496
 Db 790 TTGAAGAGCTACTGCGAGCTGTGTATGCCCTCCACACATATACGTGTCCATGTGAGTG 849
 Qy 497 gtaaggaacattgataccttcaaaagtgtccatgaagacaatttagtaagtgtcttcaata 556
 Db 850 AGAAGAGCTCCAGAACTTTCATCAAGAGCGGTCACAAAGCAATTATTTCTGTGCTCCCAATG 909
 Qy 557 ttctatgtctccaaattagaagcgtgtgaaatgatccacattccagactccagcgtgt 616
 Db 910 TCTTCTACCCAGTAACTGCTGTTCCGGTGGTGTATGCTCCCTGCTGCTCCAGGAGTGAAGTG 965
 Qy 617 attaatgtcttgcgagccttctgaagcttccatccatccagtggaataatgttataact 676
 Db 970 ACTCAACTGATGGAAGAGTCTGCTCCAGAGCTCCAGTCCGTGGAATTAATCTCTGAATA 1029
 Qy 677 tctgtggaagatttcccccgaagtcacatttgaattgtgtgcagagtgtaaaaaac 736
 Db 1030 CATGTGGAGGAGCTTCTTATTAAGAGCAATGCAAGATGCTCCAGGCTCTCAAGATGT 1089
 Qy 737 tcaatgtgaacaaatatgttggagacggtgaaaccccccaacagtaaatvgaaagattca 796
 Db 1090 TGAATGGAGGAATATGATGAGTCAAGAGTACCTCTTAAGCAAAAGAACCCGCTGGA 1149

QY 797 cttaaccatcatgaactgaacgggtgccttaataatgtgaagcttaaccataagacaa 856
 Db 1150 AATATCACTTTGAGTAGTAGAGACACATTAC-----CTAACCA 1191
 QY 857 acatcccaagaagacaccccccaatacattcagatatgttgcagtgctatttg 916
 Db 1192 ACAAGAAAGAGATGCTCCCTTATATTAATTAGATTGACAGGGAATGCGTACATG 1251
 QY 917 tttaagtcagacattgttaataatatccaacaactccatcggtcagaacttttg 976
 Db 1252 TGGCTTCCCGAATTCCTCCACACATGTTGAAGAACCTTAATCCCAACATGATG 1311
 QY 977 cctgttcaagaacacacactcctctgaagcaacttgggtcactgaatcgggttc 1036
 Db 1312 AATGGGTAAAGACACTTATAGCCAGATGAACACTGTGGGCCACCTTAGGGTGCAC 1371
 QY 1037 caagaaacacctgggagattccagatcagcccaagatg---tgtgactcgaagta 1093
 Db 1372 GGTGGATGCTGGCTCTGTCCCAACACCCCAAGTACGACATCTCAAGACATGACTTCTA 1431
 QY 1094 agactcgcctgttcaagtggaattactaagaagcttttctaccagt----- 1143
 Db 1432 TTGCCAGGCTGTCAAGTGGCAGGTCATGAGGAGACATGATAGGCTCTCTTATG 1491
 QY 1144 ----tgtactgacatcaaccttcgaagcggtgtattatgagactgcagaattaaagt 1198
 Db 1492 CTCCTGCTCTGGAATCCACACGCGGCTATCTGCGTTATGGGGCTGGGACTTGAAAT 1551
 QY 1199 ggcctatacaagaatgagactgttgttaataattgattcgaatggaacctatct 1258
 Db 1552 GGATGCTTAAACCATCATCCTGTGGCCACAAAGTTTGCCCAAGSTAGATGATATG 1611
 QY 1259 tgattaaatgcttggcagaanaagct 1283
 Db 1612 CTTTCAGTGGCTTAGAGAAATACCT 1636
 RESULT 9
 LOCUS AX045249 2216 bp DNA PAT 24-NOV-2000
 DEFINITION Sequence 1 from Patent WO0066727.
 ACCESSION AX045249
 VERSION AX045249.1 GI:11343799
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2216)
 AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
 TITLE Tumour-associated antigen
 JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
 Boehringer Ingelheim International GmbH (DE)
 FEATURES
 source
 1..2216
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..426
 427..1743
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC1734.1"
 /db_xref="GI:11343800"
 /translation="MWQKRRLCOLHYLALGCMYLATVALKLSFRKCDSDHIGLES
 RESOYCRNLIYNELKLPARKSINCSVTRGDQEAVALINLMEYKKKEPTDTH
 YLSRDEHFAERKFTOPPLSKKEVEFPYASVVIHERIENEDRLRAYAONIV
 CVHDEKPEFEKAVKALISCPVVFASLIVVAVSMERVADINCOMDILSSY
 PKRYLNTGCTDFPIKSNAMVQAKLHNGNSMSEVPRHKETRRKYHEVDL
 HLTKKQDPPTNLMTFNGNAIYASRPFVHLNFKSGLIENVKTSYEDHLVA
 TLQRRMFGVSPNHPKDYIDSMSTIALVWQVLEGDIDGATYAPGSGTHORATV
 YGAGLNMMLQNHHLANKFDPKYDNLQCLEYLRKAIYGTGL"
 1744..2216

BASE COUNT 600 a 499 c 547 g 570 t
 ORIGIN
 Query Match 14.1%; Score 191.8; DB 9; Length 2216;
 Best Local Similarity 52.5%; Pred. No. 5.7e-36;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;
 QY 317 tgaccagtgatgtgtgacattatcaagcctaagagtgatgtcgaagctgtgtcaa 376
 Db 743 TCACCAAGAGCTGTGACATTCAGAGGTGAAGAGATTCATATACGTTCCACAGCA 802
 QY 377 aggaagagaaagcttcccaatagcctatcttgggtgtccacaagaatgaattatg 436
 Db 803 AAGAAGGGTGGAGTCCCTATGCACTACCTATGATGATGATGATGATGATGATG 862
 QY 437 ttgaaggttataccatgcatatacaaccagacacatatcttagcgaatccatgtatc 496
 Db 863 TTGAAGGCTACTGCGAGCTGTATGCCCTCAGAACATATACGTGTCAATGTGATG 922
 QY 497 gtaagacactgtatcctcaagttgcataagacaacattagctaagtgtctccaata 556
 Db 923 AGAAGTCCCAAGAACTTTCAGAGGCGGTCAAGCAATATATTCTGCTCCCAAG 982
 QY 557 ttctcatgttccaatgaatgaagctgtgtgaataatgccacattccaaactccaggtc 616
 Db 983 TCTTCATAGGCACTAGACGTGGTGGGTGTTATCTCTCTGTGTCCAGGGTGAAGCTG 1042
 QY 617 attaaatgtctgtcggacacttctaagcttcaatccagttggaataatgtatcaact 676
 Db 1043 ACCCTCACTCATGGAAGACTTGTCTCCAGACGACGAGTGGCGTGAATCTTGATA 1102
 QY 677 tgtgtggcaagatttccctgaatgaatcaatttgaatgtgtcaaggttgaanaaac 736
 Db 1103 CATGTGGGAGCGACTTCCATTAAGAGCAATCAGAGATGTCAGGCTCTCAAGATGT 1162
 QY 737 tcaatgagcaaatatgttggagcgttgaaccccaacagtaaatggaagaatca 796
 Db 1163 TGAATGGAGAGATAGCATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1222
 QY 797 cttaaccatgaactgaacgggtgccttaataatgtgaagcttaaccataagacaa 856
 Db 1223 AATATCACTTTGAGTAGTAGAGACACATTAC-----CTAACCA 1264
 QY 857 acatcccaagaagacaccccccaatacattcagatatgttggcagtgctatttg 916
 Db 1265 ACAAGAAAGAGATGCTCCCTTATATTAATTAGATTGACAGGGAATGCGTACATG 1324
 QY 917 tttaagtcagacattgttaataatatccaacaactccatcggtcagaacttttg 976
 Db 1325 TGGCTTCCCGAATTCCTCCACACATGTTGAAGAACCTTAATCCCAACATGATG 1384
 QY 977 cctgttcaagaacacactcctctgaagcaacttgggtcactgaatcgggttc 1036
 Db 1385 AATGGGTAAAGACACTTATAGCCAGATGAACACTGTGGGCCACCTTAGGGTGCAC 1444
 QY 1037 caagaaacacctgggagattccaagatcagcccaagatg---tgtgactcgaagta 1093
 Db 1445 GGTGGATGCTGGCTCTGTCCCAACACCCCAAGTACGACATCTCAAGACATGACTTCTA 1504
 QY 1094 agactcgcctgttcaagtggaattactaagaagcttttctaccagt----- 1143
 Db 1505 TTGCCAGGCTGTCAAGTGGCAGGCTATGAGGAGACATGATAGGCTCTCTTATG 1564
 QY 1144 ----tgtactgacatcaaccttcgaagcggtgtattatgagactgcagaattaaagt 1198
 Db 1565 CTCCTGCTCTGGAATCCACACGCGGCTATCTGCGTTATGGGGCTGGGACTTGAAAT 1624
 QY 1199 ggcctatacaagaatgagactgttgttaataattgattcgaatggaacctatct 1258
 Db 1625 GGATGCTTCAAAACCATCATCCTGTGGCCCAAAAGTTTGCCCAAGSTAGATGATATG 1684
 QY 1259 tgattaaatgcttggcagaanaagct 1283


```

Db      1665 CTCCTCAGTGGTTGAGAGAAATACCT 1709
RESULT 10
AC022480/c
LOCUS   AC022480.5
DEFINITION Homo sapiens chromosome 15 clone RP11-361D15, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
ACCESSION AC022480.5
VERSION   GI:9230840
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS  Waterston,R.H.
TITLE    The sequence of Homo sapiens clone
JOURNAL  Unpublished
REFERENCE
AUTHORS  Waterston,R.H.
TITLE    2 (bases 1 to 181983)
JOURNAL  Direct Submission
SUBMITTED (04-FEB-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT   On Jul 16, 2000 this sequence version replaced gl:7321974.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NH0361D15
----- Summary Statistics -----
Sequencing vector: M13: 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 16456 bases at least Q40
Consensus quality: 170052 bases at least Q30
Consensus quality: 172655 bases at least Q20
Insert size: 16800; agarose-fp
Insert size: 179683; sum-of-contigs
Quality coverage: 3.76 in Q20 bases; agarose-fp
Quality coverage: 3.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1087: contig of 1087 bp in length
1088
1188
1188: gap of unknown length
2741: contig of 1554 bp in length
2742
2841: gap of unknown length
2842
4520: contig of 1679 bp in length
4521
4620: gap of unknown length
4621
6998: contig of 2378 bp in length
6999
7098: gap of unknown length
7099
10122: contig of 3024 bp in length
10123
10232: gap of unknown length
10233
14393: contig of 4171 bp in length
14394
14493: gap of unknown length
14494
18603: contig of 4110 bp in length
18604
18703: gap of unknown length
18704
22356: contig of 3653 bp in length
22357
22456: gap of unknown length
22457
26948: contig of 4492 bp in length

```

```

FEATURES
source
26949
27049
31774
31874
36374
36474
41563
41663
46646
46746
53336
53436
60197
60298
68252
68352
75761
75862
84055
84155
84156
84156
94193
94293
109723
109823
125338
125338
142873
142873
142972
142972
159683
159683
181983: contig of 22201 bp in length.
Location/Qualifiers
1. 181983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/clone="RP11-361D15"
1. 1087
/note="assembly_name:Contig19"
1188. 2741
/note="assembly_name:Contig24"
2842. 4520
/note="assembly_name:Contig25"
4621. 6998
/note="assembly_name:Contig26"
7099. 10122
/note="assembly_name:Contig27"
10223. 14393
/note="assembly_name:Contig28"
14494. 18603
/note="assembly_name:Contig29"
18704. 22356
/note="assembly_name:Contig30"
22457. 26948
/note="assembly_name:Contig31"
27049. 31773
/note="assembly_name:Contig32"
31874. 36373
/note="assembly_name:Contig33"
36474. 41562
/note="assembly_name:Contig34"
41663. 46645
/note="assembly_name:Contig35"
46746. 53335
/note="assembly_name:Contig36"
53436. 60197
/note="assembly_name:Contig37"
60298. 68252
/note="assembly_name:Contig38"
68353. 75761
/note="assembly_name:Contig39"
75862. 84055
/note="assembly_name:Contig40"
84156. 94192
misc_feature
27048: gap of unknown length
31773: contig of 4725 bp in length
31873: gap of unknown length
36373: contig of 4500 bp in length
36473: gap of unknown length
41562: contig of 5089 bp in length
41662: gap of unknown length
46645: contig of 4983 bp in length
46745: gap of unknown length
53335: contig of 6590 bp in length
53435: gap of unknown length
60197: contig of 6762 bp in length
60297: gap of unknown length
68252: contig of 7955 bp in length
68352: gap of unknown length
75761: contig of 7409 bp in length
75861: gap of unknown length
84055: contig of 8194 bp in length
84155: gap of unknown length
84156: contig of 10037 bp in length
94193: gap of unknown length
94292: contig of 15430 bp in length
109722: contig of 15430 bp in length
109822: gap of unknown length
125337: contig of 15415 bp in length
125337: gap of unknown length
142872: contig of 17335 bp in length
142872: gap of unknown length
142971: gap of unknown length
142971: contig of 16710 bp in length
159682: contig of 16710 bp in length
159683: gap of unknown length
181983: contig of 22201 bp in length.

```


Query Match 14.1%; Score 191.6; DB 58; Length 2017;
 Best Local Similarity 52.5%; Pred. No. 7.4e-36;
 Matches 518; Conservative 0; Mismatches 429; Indels 39; Gaps 3;

QY 316 atgacagatgattgtacattatcagactcgaagattatgtcgaagctgtccca 375
 |||||
 DB 551 ATACAGGAGGAGCTGTAGCGCTTCAAGGCCCAAGAAAGTTCATCTCCACTGAGT 610
 |||||
 QY 376 aagagagaaagagctccccaatagactctcttggttgtcccaagatgcaattg 435
 |||||
 DB 611 AAGAGAGAGCTGAGCTTCCCATTTGCTCTGATGATGCTGATGATGATGATG 670
 |||||
 QY 436 attgaaagctatccatgctatatacaacagcaaatattctgcatccattatg 495
 |||||
 DB 671 TTGAAGCGGCTGCTGCGAGCGCTGTATGCCCTTCAGAAATTAAGTGTGTCACGCTGAT 730
 |||||
 QY 496 cgttaagagacctgatacccttcaagttgcatgaaatattgactgaatgtctcc 555
 |||||
 DB 731 GTGAAGTCCCCAGAGAGCTTCAAGAGCGGCTCAAGGCCATTAATTCCTCCCTCCCAAT 790
 |||||
 QY 556 atttcatgtctccaatattagagctgtggaatgccaatcccaattccagactc 615
 |||||
 DB 791 GTCTCATGAGCCAGTAAGTGTGCGCGGTGTATGCTCTCCGTCGAGAGTGAAGCT 850
 |||||
 QY 616 gatttaattgtctgtcgacctctgaaagcttcaatccagttggaataatgttaca 675
 |||||
 DB 851 GACCTAAGCTGTATGGAAGAGCTGTCTCCAGAGCTCAGTCTATGGAAGTACTTACTGANT 910
 |||||
 QY 676 ttctgtgaggaagatttccctcgaagatcaatttgaattgtgtcagagttgaaa 735
 |||||
 DB 911 AACTGGGGAGAGACTTCCCATTAAGACCAATGCGAGATGATGCTGCTGCTCAAGATG 970
 |||||
 QY 736 ctcaatgagcaaatatgttggagacgtgtgaaccccccaacagtaatttgaaga 795
 |||||
 DB 971 TTGAAGGTAAGAGAGTATGAGTGTGAGTACCTTCTGAGTCCAAAAAATCGCTGG 1030
 |||||
 QY 796 acttaccatgaactgaagagcggtgcttgaatgaaatgtaagctaccacaata 855
 |||||
 DB 1031 AATTAACCGCTATGAGTGAAGACAGACAGTACCTTCCAGAGCAAGATA----- 1078
 |||||
 QY 856 aacatctcgaaggaagacccccccatcaatcagatattgtgtgagctgtatttt 915
 |||||
 DB 1079 -----AAGACCCCTCCCTCGATTAATTTACCATGTTCACAGGGAATGCTATTTT 1129
 |||||
 QY 916 gtttaagtaagcatgttgaataatatatttcaacaacccaatcgltcaagactttt 975
 |||||
 DB 1130 GTGGCCCTCTGAGCCTTTTTCACATGCTTGAACAACCTTAATCCCAATATCGGTT 1189
 |||||
 QY 976 gctctgtcctaagaacacatactctcctgatatgagcaatttgggtcactgtat 1035
 |||||
 DB 1190 GAATGGGTCAAAAGACATATATACCCCCGAGAACACCTCTGGGGCCACCTTCAGCGTGT 1249
 |||||
 QY 1036 ccaggaataactctggagagattccagatccagcaagatg---tgtctgtatcgcaagt 1092
 |||||
 DB 1250 CCGTGAGTGTGCTGTCTGTCTTCTAGCCACCCAAAGTATACATTCACAGCATGATGCC 1309
 |||||
 QY 1093 aagaactgcctgtcaagtgaatttactatgaaggtctttc-----tat 1137
 |||||
 DB 1310 ATGCCCGAGGCTCGCAAGTGGAGTACACAGAGGAGATGTTCAGCATGAGGGGGCGCTTAT 1369
 |||||
 QY 1138 cccagttgtatgactcctcagcttggaaagtggttattatgaaatgcaatgaatt 1197
 |||||
 DB 1370 GCACCCCTGCTCTGAAATCCATCGAGAGGCGCATCTGCATTTACGGGGCGGGAGCTGTAC 1429
 |||||
 QY 1198 tgggttcaaaagatgagacttgggttgcataaatttgaattcctaagtgagccatc 1257
 |||||
 DB 1430 TGAATCTCCAGAACATCACTCTTGTGCAAAAGATTGACCCGAGGGGTGATGATTAAC 1489
 |||||
 QY 1258 ttgattaaatgcttggcagaagaagt 1283
 |||||
 DB 1490 GTCTGAGTGTCTTGAAGAGTACTT 1515
 |||||

RESULT 12
 MMU68182
 LOCUS
 DEFINITION
 Mus musculus large I antigen-forming
 beta-1,6-N-acetylglucosaminyltransferase mRNA, complete cds.
 068182.1 GI:1763028

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE

Magnet, A.D. and Fukuda, M.
 Expression of large I antigen forming
 beta-1,6-N-acetylglucosaminyltransferase in various tissues of
 adult and embryonic mice
 Unpublished
 2 (bases 1 to 2393)

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Magnet, A.D. and Fukuda, M.
 Direct Submission
 Submitted (27-AUG-1996) Glycobiology Program, The Burnham
 Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
 source
 1..2393
 /organism="Mus musculus"
 /db_xref="taxon:10090"

CDS
 237..1439
 /codon_start=1
 /product="large I antigen-forming
 beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="AAB39621.1"
 /db_xref="GI:1763028"
 /translation="MPLSVRFVIVTVMVCVSVYLSFSGDQSYOKLNTSDSVRL
 SOYCSFTRDKSRFLRNKIMHEKSCFXYVQSHYITAPISOEGDPLAVYWIH
 HNDTFARLRLFMPTNITGVHDEKATREFGAYBOLVSCPYNFLASKMEPVYVG
 GISRLODLNCLIDLSTSEVPWYVAITTCQDPLPTNKEIVGLGSKNTPYL
 PPAHILRLRYVHREHLSKELSYVIRTLAKPPPNLTVIFSAVALVSRANVL
 RDPRAVDLHMSKDTFSPDEHFVTLNRIPIGVGSPNMSMTGNLNAVMMDEAHG
 GCGHYVHGICIVGNGLQWLINSOSLFANKFELNLYPIVVECTELRLRERLINOSES
 ALPSWYF"

BASE COUNT 610 a 599 c 583 g 601 t
 ORIGIN

Query Match 14.0%; Score 191.2; DB 94; Length 2393;
 Best Local Similarity 52.6%; Pred. No. 9.3e-36;
 Matches 502; Conservative 0; Mismatches 428; Indels 24; Gaps 3;

QY 337 tatcagactcgaaggttatgtcacaagcttgtctcaaggaggaagaaagcttccca 396
 |||||
 DB 459 TATGTGACCCAAAGCCACTATATCACTGCCCTTATCTCAGAGAGAGGATTTTCT 518
 |||||
 QY 397 atagcctattcttggttgtccacaagaatgcaattatggttgaaggcttatccatg 456
 |||||
 DB 519 TTGGCGTATGATGTGATCATCATCAATTGTGACACTTTCGAAGGCTCTTCAGGGCA 578
 |||||
 QY 457 atatacaaccagacaataattactgcatcatttgttgcgtaagagcccgatactc 516
 |||||
 DB 579 ATCTTCATGCCCCTCAAAATATCTATCTGTCACGTGATGAAGAAAGCAACAGCTGAATTC 638
 |||||
 QY 517 aagattgcaatgaacaattatgactgaagtgctctccaaatatttcaatgtctccaaat 576
 |||||
 DB 639 AAGAGTCCGTGGAACAGTATAGTACGCTGCTCCCAATATGCTTCTGCTTCAAGATG 698
 |||||
 QY 577 gtagctgtggaatgacacattccagactcgaagctgtatcaattgcttgcgac 636
 |||||
 DB 699 GAGCGGTGTGCTATGATGTGGAATCTCCGGCTCCAGGGGTGACCTTAAGTCAAAAGAC 758
 |||||
 QY 637 ctctcgaagcttcaactcagtggaataatgttcaactgtgtggtgcaagatttccc 696
 |||||
 DB 759 CTTTCACCTCCGAGGTCCCTTGGAATGACCATCAACACTCTGTGACAAAGACTTCCC 818
 |||||

```

QY 697 ctgaagcacaatttgatctgtgtcagagctgaacaaatcgaagcaatggttg 756
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 819 CTGAACCAACAGAGAGTAGTCTACTACTGAAAGGCTTTAAGGGAGAACCTACT 878
QY 757 gagaagtgaaacccccaaacaataattgaaagatctaccatctgtaactaga 816
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 879 CCCGGGTGCTGCTCCAGCCACGCCATTGGAAGACAGGTACGTCCA-----CCGG 932
QY 817 cgggtgcttatgataatgtgaagctaccataaagacaacatctccaaggaagacc 876
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 933 GAACACTAGCAAGAGCTTCTCTAGCTGATCAGAACACAGGGCTGAGACCCCACT 992
QY 877 ccccaataacatcagataattgttggcagtgctattgttttaagtaagaacattgt 936
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 993 CCCACACACCTCACCATTTATTTTGCTGCTCCATGCTCCCGCTGAGAGAGATTGCT 1052
QY 937 aaatatttttcaacaactccatcgttcaagacttttgcgtgtcctaagacacatac 996
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1053 AACTTGTCTCCGTGACCCAGCGCGGTGATTGCTCCATTGCTCAANAGATACCTTC 1112
QY 997 tctctgaagacacatttgggtacctgattcgggttccaggaataacctggagagatt 1056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1113 AGTCCCGATGACATTCTGGTGACACTCAATAGATTCAGAGAGTCCCTGGCTGCCA 1172
QY 1057 tccagatcagccagagatgtgtctgactcagagtaagaactcgcctgtcgaagtgaat 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1173 CCAAGCATCTCTGAGCG-----GTAACTCCAGAGCTGTGAAATGATG 1217
QY 1117 tacttaagagcttttctatccagtgactgtactcacccttcgaagcgtgtgact 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1218 GACATGGAAG---CGAAGCATGAGAGGCTGCCAGAGTACATGCTCATGGAATTGTATC 1274
QY 1177 tatggagctgaagaatgaagtggtctatcaagaatgagacttggttgcataaattt 1226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1275 TATGGAACGAGAGCTTCTCACTGCTGATTATTCGCAAAACCTGTTTCTCAAAATTT 1354
QY 1237 gatttaagtgagaccctatcttgaataatgcttgagcaaaaagccttgaagaa 1290
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1335 GAATCAACACATACCTCTTACCGTGAATGCTCGAATGAGAGCTTCGAGAA 1388

RESULT 13
AL139039 87636 bp DNA PRI 30-NOV-2000
LOCUS Human DNA sequence from clone RP11-360019 on chromosome 6, complete
DEFINITION sequence.
ACCESSION AL139039
VERSION AL139039.17 GI:11544444
KEYWORDS HTG.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 87636)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Dec 4, 2000 this sequence version replaced gi:11322778.
requesters: clonerequests@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession

```

numbers given in the feature table with their source databases:
 Emi, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP11-360019 is from the library RPCT-11.2 constructed at the
 Roswell Park Cancer Institute by the group of Pieter de Jong. For
 further details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-360019 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-421M1 is at 87537 in this sequence.
 The true right end of clone RP1-290110 is at 99 in this sequence.
 Location/Qualifiers

```

source
1..87636
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-360019"
/clone_lib="RPCT-11.2"
1..387
repeat_region
/note="MLT1B repeat: matches 1..387 of consensus"
1036..1404
repeat_region
/note="MER1B repeat: matches 2..336 of consensus"
1813..2111
repeat_region
/note="AluX repeat: matches 1..299 of consensus"
2354..2465
repeat_region
/note="MIR repeat: matches 21..143 of consensus"
2492..2690
repeat_region
/note="GPRS repeat: matches 792..981 of consensus"
2833..3354
repeat_region
/note="LIR12 repeat: matches 1..427 of consensus"
3462..3781
repeat_region
/note="LIR12 repeat: matches 80..647 of consensus"
3783..3938
repeat_region
/note="AluSg repeat: matches 134..298 of consensus"
3955..4226
repeat_region
/note="AluSg repeat: matches 1..274 of consensus"
4229..4352
repeat_region
/note="AluSg/FLAM repeat: matches 3..131 of consensus"
4714..4996
repeat_region
/note="AluX repeat: matches 1..282 of consensus"
6304..6357
repeat_region
/note="27 copies 2 mer gt 74% conserved"
6516..6817
repeat_region
/note="AluSg repeat: matches 1..303 of consensus"
7584..8250
repeat_region
/note="L1MD3 repeat: matches 7076..7738 of consensus"
8506..8649
repeat_region
/note="MIR repeat: matches 112..258 of consensus"
8679..8979
repeat_region
/note="AluSg repeat: matches 1..304 of consensus"
9000..9236
repeat_region
/note="MIR repeat: matches 17..262 of consensus"
9357..9665
repeat_region
/note="AluX repeat: matches 1..309 of consensus"
9704..10003
repeat_region
/note="AluX repeat: matches 1..297 of consensus"
10083..10505
repeat_region
/note="L1ME2 repeat: matches 5736..6158 of consensus"
10506..10766
repeat_region
/note="AluSg repeat: matches 2..297 of consensus"
10767..10860
repeat_region
/note="L1ME2 repeat: matches 5643..5736 of consensus"
10862..11056
repeat_region
/note="MER82 repeat: matches 444..652 of consensus"
11094..11400
repeat_region
/note="AluX repeat: matches 1..304 of consensus"
11522..11570
repeat_region

```


Db 102271 CACAGTGACATCCAAAGAGCTTGGCAGCTTTGAGAGGCTCTTCAGGGCGATTATAT 102330
 Qy 465 ccaagacatattacgacatccatattatgacgaagcccttacccttcaagctgc 524
 Db 102331 GCGCCAAATGCTTACTGTGTGACCTGGATGACAGGCGAGGATGCTTTAAAGTGC 102390
 Qy 525 catgaacatttagtaagtgtcttcccaatttcaatttcttcccaatttagtgct 584
 Db 102391 AGTGAACAGTTACTGACCTGCTCCCAATGCTTTCTGCTCCCAAGAGAGTGGCT 102450
 Qy 585 ggaatattgcccacatttccacatccagcctgatttaattgtctgctgcagacctcga 644
 Db 102451 TGTCTATGGGGGATCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 102510
 Qy 645 gtcttcaatcagtggaatatgtatcaattgtgtggaagatttccctcgaagtc 704
 Db 102511 CTCGTAATTCCTGGAGTATGTCATCAACCTGGGGGACAGACTTTCCCTGAAAC 102570
 Qy 705 aaatttgattgtgtcagaggttgaaacacacacacacacacacacacacacacac 764
 Db 102571 CACAGGGAATAGTTCAGTATCTGAAAGGATTTAAAGGAAATATACACCCCGAGT 102630
 Qy 765 gaaccccccaacagtaatttggaagattcacttaccatgacatgaagcaggtgc 824
 Db 102631 GCTGCTCTCTGACACAGCTGTTGGACGACATTAATACCTCCACCA-----AGAACTGTT 102684
 Qy 825 ttaagaatatggaagtaacacataagacaacacacacacacacacacacacacacac 884
 Db 102685 AATCCACAAATTCCTGATCTGAAATTAACAAATTAATTAATTAATTAATTAATTA 102744
 Qy 885 cattcaagattgtgtggaagtgcttattgttttaagtcagcagattgttgaatatat 944
 Db 102745 CATGGTATTTCTTGTGACAGGCTTACGTGCTCTCAAGAGACTTTGCTAATTGCT 102804
 Qy 945 ttccacaactccatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1004
 Db 102805 CCTCCAGACGACGCTGACGCTGATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 102864
 Qy 1005 tgaagcatttgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1043
 Db 102865 CGAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102903

RESULT 15
 AC021626/c
 LOCUS
 DEFINITION Homo sapiens clone RP11-391I2, WORKING DRAFT SEQUENCE, 28 unordered
 pieces
 AC021626
 VERSION AC021626.3 GI:7229839
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 196788)
 Homo sapiens, clone RP11-391I2
 Unpublished
 2 (bases 1 to 196788)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Balwit, J., Barna, N., Beckler, R., Beda, F.,
 Bogenlasky, L., Bouknight, B., Brown, A., Burkett, G., Castle, A.,
 Chospel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Deavello, K., Dewar, K., Domino, M., Doyle, M., Fennell, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howard, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
 Landers, T., Lebeck, J., Levine, R., Liu, G., Locke, K.,
 Macdonald, P., Margulis, N., McEwen, P., McGurk, A., McKernan, K.,
 McPheters, R., Meldrum, D., Meneses, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,

TITLE JOURNAL COMMENT

Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A., and Zody, M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 12, 2000 this sequence version replaced g1:6978227.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 15883
 Center clone name: 391.I-2
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 181094 bases at least Q40
 Consensus quality: 189581 bases at least Q30
 Consensus quality: 192532 bases at least Q20
 Insert size: 208000; agarose-fp
 Insert size: 194088; sum-of-contigs
 Quality coverage: 3.7 in Q20 bases; agarose-fp
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1281: contig of 1281 bp in length
 * 1282 1381: gap of 100 bp
 * 1382 2889: contig of 1518 bp in length
 * 2890 2999: gap of 100 bp
 * 3000 5681: contig of 2682 bp in length
 * 5682 5781: gap of 100 bp
 * 5782 8351: contig of 2570 bp in length
 * 8352 8451: gap of 100 bp
 * 8452 10639: contig of 2188 bp in length
 * 10640 10739: gap of 100 bp
 * 10740 13653: contig of 2916 bp in length
 * 13656 13755: gap of 100 bp
 * 13756 15927: contig of 2172 bp in length
 * 15928 16027: gap of 100 bp
 * 16028 19028: contig of 3001 bp in length
 * 19029 19128: gap of 100 bp
 * 19129 22669: contig of 3541 bp in length
 * 22670 22769: gap of 100 bp
 * 22770 26791: contig of 4022 bp in length
 * 26792 26891: gap of 100 bp
 * 26892 30845: contig of 3954 bp in length
 * 30846 30945: gap of 100 bp
 * 30946 35594: contig of 4649 bp in length
 * 35595 35694: gap of 100 bp
 * 35695 40148: contig of 4454 bp in length
 * 40149 40248: gap of 100 bp
 * 40249 45042: contig of 4794 bp in length
 * 45043 45142: gap of 100 bp
 * 45143 49477: contig of 4335 bp in length
 * 49478 49577: gap of 100 bp
 * 49578 54518: contig of 4941 bp in length
 * 54519 54618: gap of 100 bp
 * 54619 61983: contig of 7364 bp in length
 * 61983 62082: gap of 100 bp


```

* 62083 67811: contig of 5729 bp in length
* 67812 67911: gap of 100 bp
* 67912 74070: contig of 6159 bp in length
* 74071 74170: gap of 100 bp
* 74171 80685: contig of 6515 bp in length
* 80686 80785: gap of 100 bp
* 80786 89056: contig of 8271 bp in length
* 89057 89156: gap of 100 bp
* 89157 99104: contig of 9948 bp in length
* 99105 99204: gap of 100 bp
* 99205 109014: contig of 9810 bp in length
* 109015 109114: gap of 100 bp
* 109115 132471: contig of 13357 bp in length
* 132472 132571: gap of 100 bp
* 132572 138753: contig of 16182 bp in length
* 138754 138853: gap of 100 bp
* 138854 156691: contig of 17838 bp in length
* 156692 156791: gap of 100 bp
* 156792 173620: contig of 16829 bp in length
* 173621 173720: gap of 100 bp
* 173721 196788: contig of 23068 bp in length.

```

FEATURES

```

Source
1. 196788
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="RP11-391I2"
   /clone_1b="RP11-391I2"
misc_feature
1. 1281
   /note="assembly_fragment"
misc_feature
1382..2899
   /note="assembly_fragment"
misc_feature
3000..5681
   /note="assembly_fragment"
misc_feature
5782..8351
   /note="assembly_fragment"
misc_feature
8452..10639
   /note="assembly_fragment"
misc_feature
10740..13655
   /note="assembly_fragment"
misc_feature
13756..15927
   /note="assembly_fragment"
misc_feature
clone_end:T7
vector_side:right
16028..19028
   /note="assembly_fragment"
misc_feature
19129..22669
   /note="assembly_fragment"
misc_feature
22770..26791
   /note="assembly_fragment"
misc_feature
26892..30845
   /note="assembly_fragment"
misc_feature
30946..33594
   /note="assembly_fragment"
misc_feature
33695..40148
   /note="assembly_fragment"
misc_feature
40249..45042
   /note="assembly_fragment"
misc_feature
45143..49477
   /note="assembly_fragment"
misc_feature
49578..54518
   /note="assembly_fragment"
misc_feature
54619..61982
   /note="assembly_fragment"
misc_feature
62083..67811
   /note="assembly_fragment"
misc_feature
67912..74070
   /note="assembly_fragment"
misc_feature
74171..80685
   /note="assembly_fragment"
misc_feature
80786..89056
   /note="assembly_fragment"
misc_feature
89157..99104
   /note="assembly_fragment"

```

```

misc_feature 99205..109014
   /note="assembly_fragment"
misc_feature 109115..132471
   /note="assembly_fragment"
misc_feature 132572..138753
   /note="assembly_fragment"
misc_feature 138854..156691
   /note="assembly_fragment"
misc_feature 156792..173620
   /note="assembly_fragment"
misc_feature clone_end:SP6
vector_side:right
misc_feature 173721..196788
   /note="assembly_fragment"
BASE COUNT 53672 a 43305 c 43403 g 53703 t 2705 others
ORIGIN

```

```

Query Match 13.8%; Score 187.4; DB 66; Length 196788;
Best Local Similarity 55.4%; Pred. No. 1e-34;
Matches 387; Conservative 0; Mismatches 306; Indels 6; Gaps 1;

QY 345 tcctgaaggtatgctcaaaagctgtgtccaaaggaggaagaaagctcccaatagccta 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78350 TCGAAGCCACATGTGTACAGAAACACTCTGTAAGAAAGGCTGGCTTCCCTTAGCTTA 78291

QY 405 tctctgtgtgtcccaaaagatgcaattatggttgaaaggcttatccatgtatataca 464
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78290 CACAGTGACCATTCACAAAGACTTCGGCATTTTGAAGGCTCTTCAGGGGATTATAT 78231

QY 465 ccagacacatatctatcgtatccattatgcatgtaaggacacctgatactccaagtgc 524
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78230 GCCCCAAATGTCCTACTGTGTGTCACCTGTGATCAGAACGACGAGATGCTTTAAAGTGC 78171

QY 525 catgacaattagctaaagctcttccaataatttcaatttcaatttcaatttcaatttca 584
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78170 AATGAAACAGTACCTACAGTGTCTCCCAATGCTTTGTGCTCCCAAGAAAGAGTCCGT 78111

QY 585 ggaataatgccacattccagacatccagacatccagacatccagacatccagacatccag 644
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78110 TGTCTATGGGGGATCTCCAGGCTCCAGGCTCCAGGCTCCAGGCTCCAGGCTCCAGG 78051

QY 645 gttctcaatccagtggaatatgatactatgctgtgtgtgtgtgtgtgtgtgtgtgtgt 704
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78050 CTTGTAAAGTTCCTGGAATATGTCATCAACCTCGCGGCAAGACTTCCCTGTAATAC 77991

QY 705 aaatttgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 764
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77990 CAACAGGAAATAGTTCATCTATCTGAAGGATTAAAGGAAATATACCCCGGAGT 77931

QY 765 gaaaccccaacagtaaatgtgaaagattcaattacatcatgaaacttaagcgggtgcc 824
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77930 GGTGCTCTCTGACACGCTGTGAGAGGACTTAATACGTCCCA-----AGAACTGT 77877

QY 825 ttatgatatgtgaaagctaacaaaggaagaaacatcccaaggaagaccccccataa 884
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77876 AAACCACAAAATTCCTACGTATTAACAAACAAATTAATAACTCTCTCTCAATGA 77817

QY 885 catcagatattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 944
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77816 CATGGAGATTACTTGTGGACGCGCTACGTGCTCTCAAGGAGACTTGGTACTTGT 77757

QY 945 tttaaacatccatcagcttcaagactttttgctgtgtgtgtgtgtgtgtgtgtgtgtgt 1004
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77756 CCTCCAGACGCTGCGACTTGACTTACTCTCTGCTGCAAGGACCTTACAGCCCGA 77697

QY 1005 tgaacactttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1043
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77696 CGAATCTTCTGGTGACACTCAACAGGATTCCTCCGTAT 77658

```


Fri Aug 3 09:38:03 2001

Job time: 6969 sec

us-09-645-292-1.rge

Page 19


```

Db 159 IYCHDRKAPDTFKYAMNNLAKCSNIFIASKLEAVEYAHISRLOADNCLSDLKSSI 218
QY 181 QMKYVNLGCGDPLKSNFELVSELKKNANLLEVPKNSKLEFTHHELRVPY 240
Db 219 QMKYVNLGCGDPLKSNFELVSELKKNANLLEVPKNSKLEFTHHELRVPY 278
QY 241 VKIPRTNSKEAPHNIQIFVGSAYFVLSQAFVKYIFNNSIVQDFEAMSKDTYSPDEH 300
Db 279 VKIPRTNSKEAPHNIQIFVGSAYFVLSQAFVKYIFNNSIVQDFEAMSKDTYSPDEH 338
QY 301 WAILIRPQIPGELSRACQVSDQSKTRLYKKNYVEGFYPSCTSHLRVCITYAAEL 360
Db 339 WAILIRPQIPGELSRACQVSDQSKTRLYKKNYVEGFYPSCTSHLRVCITYAAEL 398
QY 361 WMLIKDGHMFANKFDSKVPDILIKCLAEKLEEQORDMITLPESEKLEMDRNLTTTS 415
Db 399 WMLIKDGHMFANKFDSKVPDILIKCLAEKLEEQORDMITLPESEKLEMDRNLTTTS 453

RESULT 2
ID 095395 PRELIMINARY; PRT: 438 AA.
AC 095395;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE.
GN C2/4GNT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99115671; PubMed=9915862;
RA Yeh J.C., Ong E., Fukuda M.;
RT "Molecular cloning and expression of a novel beta-1,6-N-
RT acetylglucosaminyltransferase that forms core 2, core 4, and I
RT branches."
RL J. Biol. Chem. 274:3215-3221(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99143102; PubMed=9988682;
RA Schiavetek T., Nomoto M., Levery S.B., Merx G., van Kessel A.G.,
RA Bennett E.P., Hollingsworth M.A., Clausen H.;
RT "Control of O-glycan branch formation. Molecular cloning of human cDNA
RT encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core
RT 2 and core 4."
RL J. Biol. Chem. 274:4504-4512(1999).
DR EMBL; AF102542; AAD10824.1;
DR EMBL; AF038650; AAD21525.1;
DR InterPro; IPR003406;
DR Pfam; PF02485; Branch; 1.
DR Transferrase; Glycosyltransferase.
SQ SEQUENCE 438 AA: 50863 MW: 1FF0A7B451C88407 CRC64;

Query Match 39.3%; Score 862; DB 4; Length 438;
Best Local Similarity 44.5%; Pred. No. 1.5e-59;
Matches 173; Conservative 70; Mismatches 128; Indels 18; Gaps 8;

```

```

Db 166 EKSPTEKAVKAILISCPNVLASKLVRVYASWRSQVADNCEMLLOSSVPMTYFN 225
QY 188 LOGDPLKSNFELVSELKKNANLLEVPKNSKLEFTHHELRVPY 247
Db 226 TCGDTFPIKSNAMVQALKMLNGRNSMESEVPKHEIRMKYHEVYR---DTLHL---T 279
QY 248 NISKAPPHNIQIFVGSAYFVLSQAFVKYIFNNSIVQDFEAMSKDTYSPDEHFNATLIRV 307
Db 280 NKKDDPPYNTLMFIGNAYIVASRDEVOHLKNSQDLLEWVDYTSPEHLATLQRA 339
QY 308 PGIPGEI-SRSADQVSDQSKTRLYKKNYVEGF-----YPSCTSHLRVCITYAAELR 361
Db 340 RMFGSVNHRKPYDLSMDTSLARLYKMGHGDDIDKAPAPACGIGHORALCYGAGDNLN 399
QY 362 WMLKDHMFANKFDSKVPDILIKCLAEKL 390
Db 400 WMLQNHHLANKFDPKVDNALQCLEEYL 428

RESULT 3
ID 064165 PRELIMINARY; PRT: 428 AA.
AC 064165;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE ENZYMAIC GLYCOSYLATION-REGULATING GENE.PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;
RX MEDLINE=96013636; PubMed=7560067;
RA Nishio Y., Warren C.E., Buczek-Thomas J.A., Rulfs J., Koya D.,
RA Aiello L.P., Feener E.P., Miller T.B., Jr., Dennis J.W., King G.L.;
RT "Identification and characterization of a gene regulating enzymatic
RT glycosylation which is induced by diabetes and hyperglycemia"
RT specifically in rat cardiac tissue."
RL J. Clin. Invest. 96:1759-1767(1995).
DR EMBL; S79797; AAB35697.2;
DR InterPro; IPR003406;
DR Pfam; PF02485; Branch; 1.
SQ SEQUENCE 428 AA: 49826 MW: AE25A18172897AA8 CRC64;

Query Match 38.9%; Score 854.5; DB 11; Length 428;
Best Local Similarity 42.1%; Pred. No. 5.5e-59;
Matches 177; Conservative 71; Mismatches 133; Indels 39; Gaps 10;

```

Df				302	NKNIQFMWAOQTSPDEPLMTATIQRIPEVPSGLSPSSHEKYDLSDMNNAVARVMQPVEG	36d
Oy	339	F-----YPCSTGSHLRSVCIVTGAALRLNLKDGHFFAKKEPSSKVDPILICLTAKELEO		393		
		: : : : :				:
Db	362	DVSNGAPTPCSCGVHVRASCVCFVGSLXWLKRHHFPAKKFDMDVDYPALOCLBEBLRHK		42t		
RESULT	4					
OSI7KZ		PRELIMINARY;	PRT:	440 AA.		
OSI7KZ						
ID	OSI7KZ;					
AC	OSI7KZ;					
DT	01-OCT-2000 (TREMBLrel; 15,	Created)				
DI	01-Oct-2000 (TREMBLrel; 15,	last sequence update)				
DT	01-MAR-2001 (TREMBLrel; 16,	last annotation update)				
De	BETA-1,-N-ACETYLGUCOSAMINYLTRANSFERASE.					
Gn	BOFEF-4.					
Os	Bovine herpesvirus 4.					
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;					
OC	Gammaherpesvirinae.					
Ox	NCBI_taxid=10385;					
Rn	[1]					
Rp	SEQUENCE FROM N.A.					
Rr	STRAIN=V. TEST;					
Rv	MEDLINE+20283898; PubMed-10811884;					
Rz	Vanderplasseben A., Markne-Chlapphoff N., Lomonte P., Suzuki M.,					
Rz	Hiroaka N., Yen J.-C., Bureau F., Willems L., Thiry E., Fukuda M.,					
Rz	Pastoret P.-P.;					
Rz	"A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded					
Rt	by Bovine herpesvirus type 4."					
Rt	Proc. Natl. Acad. Sci. U.S.A. 97:5756-5761(2000).					
Rt	EMBL: AF31105; AAF72001.1; "					
Dr	InterPro: IPR003406; "-"					
KR	pfam: PF02485; Branch: 1					
DR	(transferase; Glycosyltransferase.					
Sq	SEQUENCE 440 AA; 50701 MW; 71B9C3B6BA949BD CRC64;					

[illegible]

AC 035981; Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, last sequence update)
 DT 01-JAN-1998 (TrEMBLrel. 05, last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, last annotation update)
 DE GLUCOSAMINYL (N-ACETYL) TRANSFERASE 1, CORE 2 (BETA-1,6-N-
 DE ACETYLGLUCOSAMINYLTRANSFERASE).
 GN GCNT1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=BALB/C; TISSUE=SUBMAXILLARY GLAND, AND KIDNEY;
 RC MEDLINE=98001705; PubMed=9341170;
 RA Sekine M., Nara K., Suzuki A.;
 RT "Tissue-specific regulation of mouse core 2 beta-1,6-N-
 RT acetylglucosaminyltransferase.";
 RL J. Biol. Chem. 272:27246-27252(1997).
 RL EMBL; D87333; BAA2899.1; -.
 DR EMBL; D87333; BAA2899.1; -.
 DR MGD; MGI:95676; Gcnrl.
 DR InterPro: IPR003406; -.
 DR Pfam: PF02485; Branch: 1.
 KM Transferrase, Glycosyltransferase.
 SQ SEQUENCE 428 AA; 49851 KM; F0D10A6ADB23B92C CRC64;

[illegible]


```

Db 127 QPONOFLAVDGNSSVEFGLVRLMRCYGNIOYFTIDEIRWCGEILTSV---FOCVDY 183
OY 175 LKSSIQMKYYINLCGDFPLKSNPELVSELKILNGANMLTVPKPNKSLERFYHHLR 234
Db 184 LAKLPDMKRYOYISGVADPAKSNLEMTIRIKALNG-----SENKEI- 225
OY 235 RVPYEVKLPRTNISKAPPHNIOIFVGSAYFVLQAQVYIFNNSIVODEFAMSKDTY 294
Db 226 -LPFEFYRLNKRKPMWSSPLPKYKSL---SATF---SRKSANFVNSEKVLQIDFLRGTT 279
OY 295 SPDEFKATILRVPCI-----PGEISRSADQVSDL-QSKTLVYK-----NYE 337
Db 280 CADESLMKTIGNPKYKFSLEPMGFDAKAMHKNNRTGKLGKXGENOKIDNGYV 339
OY 338 GFFYP-----SCTGSHLRVCIYGAELRMLIKDGHMFANKFDSKYDPIILKLAETL 390
Db 340 SRQOYVNRAPYKCKGYRSLSCVGVYDLPNLINRHELVAKLYFSYQPAAFCLVENS 399
OY 391 EQQ 393
Db 400 ROK 402

```

```

RESULT 9
002314 PRELIMINARY; PRT: 401 AA.
AC 002314
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE T15D6.2 PROTEIN.
GN T15D6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL: 283125; CAB05620.1; -
DR InterPro: IPR003406; -
DR Pfam: PF02485; Branch: 1.
SQ SEQUENCE 401 AA; 45608 MW; EID3F8AD84C0873D CRC64;

```

```

Query Match 13.5%; Score 296; DB 5; Length 401;
Best Local Similarity 23.4%; Pred. No. 2.7e-15;
Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;
OY 23 THKDEVRVNCSCGYEOPLEIGKSLERIRRDITLDEDDVAVMTSCDIYQTLRGYA 82
Db 34 TDLDDL--EINCNIL--QGLKNEKQELINTKMI---EDKIMNSTDCCHVYKSMFRN 86
OY 83 QKTVSEKSEFPYLAISLVYHKDAIMVERLIHALYNOHNITCYHIDKRAKADPTKVAAMNIA 142
Db 87 EYPLSEEARFPLSYGLVYKELSOVLFMILSSLYOPQNEYCIAVGENSASTLILLELS 146

```

```

OY 143 KCFSNIFIAKLEAVEYAHISRLAOLNCLSLKSSIQMKYYINLCGDFPLKSNPELV 202
Db 147 DCFPN-----KRPITMGSEILINSYDDCKFLSHLKSMMKRYEQYISGVDIPLKTLNAV 201
OY 203 SELKILNGANMLTVPKPNKSLERFYHHLRVPEYVKLPRTNISKAPPHNIOITV 262
Db 202 RLKSLNGTANVEIKYENRIL-----LGNETESPFLK 237
OY 263 GSAVFLVQAQVYIFNNSIVODEFAMSKDTYSPDEFKATILRVPG---IPGEISRSQA 319
Db 238 SSLSLIPKRAANYIASSSIPOQLLEFRNTWADGFCNGTLFDVGSNL----- 293
OY 320 DVSDLSKTRLVKMYEGEFPSCGSHLRVCIYGAELRMLIKDGHMFANKFDSKYD 379
Db 294 -----FEEHQIFESGCHNHMKDSCVFGIDVSNLLOAKALVAKLYLISE 340
OY 380 PILIKL 386
Db 341 PEAFVCL 347

```

```

RESULT 10
019729 PRELIMINARY; PRT: 454 AA.
ID 019729
AC 019729;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE F22D6.12 PROTEIN.
GN F22D6.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL: 271262; CA95816.1; -
DR InterPro: IPR003406; -
DR Pfam: PF02485; Branch: 1.
SQ SEQUENCE 454 AA; 52822 MW; FA0134DBF31D0925 CRC64;

```

```

Query Match 13.3%; Score 291; DB 5; Length 454;
Best Local Similarity 24.1%; Pred. No. 7.9e-15;
Matches 99; Conservative 73; Mismatches 174; Indels 64; Gaps 14;
OY 17 FVRNRYTHVDEV---RYVNCSCGYEOPLEIGKSLERIRRDITLDEDDVAVMTSDC 73
Db 40 FLKSICGYAQLVSEKSEFPYLAISLVYHKDA-----IMVERLIHAI 115
OY 74 IYQTLFRNTNPMASAEKHPFLSYGLVYKDLPOVTPAKMFIKHLFLEINQVFLLSI 154
Db 95 TUKTLFRNTNPMASAEKHPFLSYGLVYKDLPOVTPAKMFIKHLFLEINQVFLLSI 154

```

[illegible]

```

RESULT 11
020406 PRELIMINARY; PRT: 472 AA.
ID Q20406;
AC Q20406;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE F44F4.6 PROTEIN.
GN F44F4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae.
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Coles L.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A
RA Craxton M., Dear S., Du Z., Dublin R., Faveilo A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jaffer M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Lettelle P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shewkeen B
RA Smalton N., Smith A., Sonnhammer E., Staden A., Sulston J.,
RA Waterson A., Weisstock L., Wilkinson-Spoot J., Wohldman P.,
RA Watson A., Weisstock L., Wilkinson-Spoot J., Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
DR EMBL: Z37092; CAA85457.1; -
DR InterPro: IPR003406; -
DR Pfam: PF02485; Barch. 1.
SQ SEQUENCE 472 AA; 5541 MW; 36A18D96D9C6245A CRC64;

```

[illegible]

```

Db      183  HAFIG-----EESQWSEFGLKLVVYCFPMILSKSKOKWVYQYLSTGDIPIRTINLEMRIK  239
Qy      207  KLNANMLEYKPPNSKLERPTYHHLRPVYEVKPIPTINISKEAPPINIOIFGSAV  266
Db      240  ALNS-----NNTDVSTF-----EVDRI-----KNNEGVLPP--MVIYKSSMS  275
Qy      267  FVLSQAEVYKIFENNSIYODFFAMSKDITVSPDEHFWALLIKVPG---IPGLISSAQVSD  323
Db      276  VVVPREGADYLISSPVOOKLTKYISKTIWIPDESSTWVLSLPALLPVGSI--RVKRIIW  333
Qy      324  LOSKTRL-----VKWVYEGF-----FYPSCTSHLSVYCIAGAEILRWLIKDGHW  369
Db      334  LRKNKLEPPYENTVNSIGTSYIRQYOWGMOKCEFGKVDSEFVEVEDIEEIMTPEL  393
Qy      370  FANFEDSVADPILIKCLAERKEEQ-----RQWITLPSEKILPMDNLT  412
Db      394  VAHKLILEFQPAALFMCMLKEVRRRSLSPDALTFASRYSQMPTVELYOGAII  446

```

RESULT	12
ID	002315
AC	002315
DT	01-JUL-1997 (TREMBLrel, 04, Created)
DT	01-JUL-1997 (TREMBLrel, 04, Last sequence update)
DT	01-MAR-2001 (TREMBLrel, 16, Last annotation update)
DE	T15D6.3. PROTEIN.
GN	T15D6.3.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC	Rhabditidae; Pterodermidae; Caenorhabditis.
OX	NCBI_TaxID=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Dobson R.;
RL	Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEBLINE=94150718; PubMed=7906398;
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berts M.,
RA	Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A
RA	Craighton M., Dear S., Du Z., Dublin R., Favello A., Fulton L.,
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L.,
RA	Jones M., Keshaw J., Kirsten J., Lalster N., Latreille P.,
RA	Lilliching J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA	Parsons L., Percy C., McMurtry A., Roopra A., Saunders S., Shownkeen
RA	Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA	Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA	Watson S.A., Weinstock L., Wilkinson-Sproat J., Woldman F.,
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT	elegans."
RL	Nature 368:32-38(1994).
DR	EMBL: 283125; CAB05621.1; -
DR	InterPro: IPR003406; -
DR	Pfam: PF02485; Branch 1.
DR	SEQUENCE 402 AA; 46135 MW; F6AD721976703307 CRC64;

[illegible]

```

Db 157 NCFPN-----KRPPIKMGSEFINSYVGCLEF.LSHLKSMDKPYQYLSGVDPILKTNLEW 211
QY 203 SELKINGAMLETVKPPNSKLEFTHHRLRPVPEVYKLPITNTSKAPPNIOIFV 262
Db 212 RILKRLNG-----TVNIGIS-----TYEDRL-----LNGKNTSP--LPLFK 247
QY 263 GSAFVLSQAFVKYIFNNSIVODEFAMSKDYPSPDEHFMAF-----LIRVPGIGEISR 316
Db 248 SLSLSLIPRKANVLSSSSVQQLLEFLRTWVADGEFGWGLFEGNKDLFVPGSFNNDP 307
QY 317 SAQVSDLOSQKRLVYK-----NYEGFFPSCGSHLSVCYICGAELRWLTKDGHMF 370
Db 308 LTNMGWNVYSRHOL--WVESECHNYMKD-----RS-CVFGIGDVPNLMKSRALV 353
QY 371 ANKFDSKVDPIILKCL 386
Db 354 AHKLYIESEPEAFCL 369

RESULT 13
C02309 PRELIMINARY: PRT; 459 AA.
ID 002309
AC 002309:
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE T09E11.9 PROTEIN.
GN T09E11.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN 11
RP SEQUENCE FROM N.A.
RL McClay K.;
RN Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Crahan M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Smailson N., Percy C., Riffen L., Roopra A., Saunders D., Showkeen R.,
RA Spaldon K., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sprat J., Wohlman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
RL EMBL: Z81147; CAB03536.1;
DR InterPro: IPR003406;
DR Pfam: PF02485; Branch: 1;
SQ SEQUENCE 459 AA; 53476 MW; A514592B10F1A8E CAC64;

Query Match 11.2%; Score 245; DB 5; Length 459;
Best Local Similarity 22.8%; Pred. No. 3.2e-11;
Matches 106; Conservative 74; Mismatches 202; Indels 82; Gaps 17;

QY 5 YLVEVSLSTSPVRRKRYHVDVRYEVNCGIY--EEPELIGKSLFIRROIIDED 62
Db 25 VYVIMBS-SKSLFGRN-----EQNSYALYNLRNENYOREALINELRDFAKIES 72
QY 63 -----DPAVAMTSDDDIYQTLNGYAKLVSEKESFP----- 94
Db 73 KIEHFYSKTYRRSEFAHDCGGLISGDKDYLOTYSG--ENRIPWENNPDMSCSAVMD 130
QY 95 -----TAYSLVYHKDAIKVRELHAIYNQHNITCYHYDKKADPTKVVAN 139
Db 131 RIIPHDIILPLKNGVAFARIYVKDYELVEKOVMSYHNPQNSCFADIKKAPTRFNQMR 190

```

```

QY 140 NLAKFSNIFIAKLEAVEVAHISRLQADLANCLSDLKSSIQMKYVINLCGDFPKSNE 199
Db 191 AMACIPNVLLPDPQEPIDSNNGHNVLAHNLRLALINNP--GMNVAMLLQNDLTKSY 249
QY 200 ELVSELKYLNGANMLETVPKPPNSKLEFTHHRLRPVPEVYKLPITNTSKAPPNIO 259
Db 250 ELEQVYEWGAGNDAVELL--PEAQRLDEENFKMD-----PSLKMFPDESKVDETIINELK 304
QY 260 IFVGSAYFVLSQAFVKYIFNNSIVODEF--AMGSDYSPDEHFMAF--LIRVPGIGEISR 316
Db 305 FSKGVQGSKRAVDVMWTRKYNLSTYIDQWQNGRGVDEMLSSIQIAFLMGCHF-- 362
QY 317 SAQVSDLOSQKRLVYK-----NYEGFFPSCGSHLR--VCYICGAELRWLTKDGHMFANKFD 375
Db 363 TDQCLKEGKEPFRMYQWNEDESY---CASKYKRVHVCILGIELRSVASFPLTMNKML 419
QY 376 SKVDPIILKCLLEEO---ORDWITLPSKLEMDRLTTTS 415
Db 420 PPFDSITIECTAELLYNRTFMQNDH---PLEEET--KKNWTVS 459

RESULT 14
C09M034 PRELIMINARY: PRT; 876 AA.
ID 09M034
AC 09M034; C09M035; Q24116;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG17771 PROTEIN.
GN CG17771 OR CG17772 OR ANON-12.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN 11
RP SEQUENCE FROM N.A.
RL Tissue-IMAGINAL DISK;
RA Bessarab D.A.; Sun H.Y.;
RN Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
[1]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Barton R.C., Korman J.R., Yandell M.P., Zhang Q., Chen L.X.,
RA Brandon R.G., Rogers J.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Adair J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Buschan D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Casady S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., May S.A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Ibegyan C.,
RA Jaitani M., Kalush F., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kraut S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mestulov G., Milshina N.V., Mobarry C., Morris J., Mosteifi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

```


Fri Aug 3 09:38:18 2001

us-09-645-192-2_copy_39_453.rpt

GenCore version 4.5
Copyright (c) 1993 - 2000 Compaq Inc.

OM protein - protein search, using sw model

Run on: August 2, 2001, 14:50:18 ; Search time 16.71 Seconds

(without alignments)
850.750 million cell updates/sec

Title: US-09-645-192-2_COPY_39_453
Perfect score: 2194
Sequence: 1 F0KDIYEVSLSTSPFYRN.....DWITLPSKLFMDNLTTS 415

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	849.5	38.7	428 1	G6NT_HUMAN 002742 homo sapien
2	837	38.1	428 1	G6NT_MOUSE 009374 mus musculu
3	831.5	37.9	427 1	G6NT_BOVIN 092180 bos taurus
4	730.5	33.3	400 1	BGIB_MOUSE 097402 mus musculu
5	711	32.4	400 1	BGIB_HUMAN 006430 homo sapien
6	711	5.3	895 1	YAS4_METJA 058454 methanococ
7	108.5	4.9	661 1	WHI3_YEAST 034761 saccharomyc
8	102.5	4.7	654 1	BFR2_HUMAN 001742 homo sapien
9	102.5	4.7	821 1	FGR2_HUMAN 021802 homo sapien
10	102.5	4.7	821 1	FGR2_MOUSE 021803 mus musculu
11	100.5	4.6	2607 1	BACB_BACLI 068007 bacillus 11
12	99.5	4.5	666 1	VOO1_VACCC 021093 vaccinia vi
13	99.5	4.5	823 1	CEK3_CHICK 018461 gallus gall
14	97.5	4.4	1314 1	TEIX_CLOTE 04958 clostridium
15	97.5	4.4	2054 1	YCF2_PINTH 041653 pinus thunb
16	97.5	4.4	440 1	GAAP_HUMAN 000531 homo sapien
17	96	4.4	1036 1	YQ6B_CAREL 092483 caenorhabd
18	95	4.4	701 1	SYGB_HELPJ 092483 heliobacte
19	94	4.3	1344 1	IF3A_MOUSE 020316 mus musculu
20	94	4.3	1675 1	CLH1_HUMAN 000610 homo sapien
21	94	4.3	1675 1	CLH_BOVIN 049591 bos taurus
22	94	4.3	1675 1	CLH_RAT 041442 rattus norv
23	93.5	4.3	503 1	SECA_PORPU 051381 porphyra pu
24	93	4.2	503 1	CPV1_SHEEP 092483 ovis aries
25	93	4.2	520 1	Y208_METJA 060270 methanococ
26	93	4.2	576 1	POLO_DROME 052304 drosophila
27	92	4.2	908 1	DPO1_BORBU 051498 borrelia bu
28	92	4.2	1103 1	CYGF_BOVIN 002740 bos taurus
29	92	4.2	1382 1	IF3A_HUMAN 001452 homo sapien
30	91.5	4.2	720 1	SPOT_MYCGE 047530 mycoplasma
31	91.5	4.2	1956 1	ATX1_PLAFA 004556 plasmodium
32	91	4.1	364 1	DF3B_MYCGE 047247 mycoplasma
33	91	4.1	1037 1	Y0J8_YEAST 012456 saccharomyc

34	91	4.1	1077 1	DN2L_HUMAN	P51530 homo sapien
35	90.5	4.1	560 1	INR1_SHEEP	Q28589 ovis aries
36	90.5	4.1	759 1	TEAL_YEAST	P47988 saccharomyc
37	90.5	4.1	850 1	CYAA_YERPE	P40127 yeastinia pe
38	90.5	4.1	936 1	MSH4_HUMAN	Q15457 homo sapien
39	90.5	4.1	1036 1	Y414_MYCGE	P47653 mycoplasma
40	90.5	4.1	2280 1	YCF2_TOBAC	P09976 nicotiana t
41	90	4.1	415 1	EP1G_YEAST	P29547 saccharomyc
42	90	4.1	503 1	CPV2_PIG	P79430 sus scrofa
43	90	4.1	605 1	GLMS_THEMA	Q95425 t glucosami
44	90	4.1	1324 1	VGL2_CYMA5	P11224 murine coro
45	90	4.1	1655 1	N188_YEAST	P52593 saccharomyc

ALIGNMENTS

RESULT 1	ID	GENE	STANDARD	PRT	AA
AC	002742	G6NT_HUMAN			
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)				
DE	(CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).				
GN	GCNT1				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-93028457; PubMed-1329093;				
RA	Bierhuizen M.F.A., Fukuda M.,				
RT	"Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6-GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen."				
RL	Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-Placenta;				
RX	MEDLINE-96078409; PubMed-7579796;				
RA	Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.,				
RT	"Genomic organization of core 2 and 1 branching beta-1,6-N-acetylglucosaminyltransferases. Implication for evolution of the beta-1,6-N-acetylglucosaminyltransferase gene family."				
RL	Glycobiology 5:417-425(1995).				
CC	- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.				
CC	- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-GLUCOSAMINYL-1,6-N-ACETYL-D-GALACTOSAMINYL-R.				
CC	- PATHWAY: GLYCOSYLATION.				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.				
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES AND MYELOID CELLS.				
CC	- SIMILARITY: TO T-BRANCHING ENZYME (IGNT).				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: M97347; AAA55919.1; -				
DR	EMBL: L41415; AAA96661.1; -				
DR	PIR: A46293; A46293.				
DR	MTM: 600391; -				
KW	Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;				

Dt	15-JUL-1999	(Rel. 38, Created)
Dt	15-JUL-1999	(Rel. 38, Last sequence update)
Dt	15-JUL-1999	(Rel. 38, Last annotation update)
Dt	N-ACETYLGLUCOSAMINIDE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE	
De	(EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)	
De	(IGT) (LARGE I ANTIGEN-FORMING BETA-1,6-N-	
De	ACETYLGALACTOSAMINYLTRANSFERASE).	
Gn	GCNT2.	
Os	Mus musculus (Mouse).	
Oc	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Oc	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
Ox	NCBI_TaxID=10090;	
Rn	[1]	
Rp	SEQUENCE FROM N.A.	
Rx	MEDLINE=97280061; PubMed=9134435;	
Ra	Magner A.D., Fukuda M.;	
Rt	"Expression of the large I antigen forming beta-1,6-N-	
Rl	acetylglucosaminyltransferase in various tissues of adult mice.";	
Rl	Glycobiology 7:285-295(1997).	
Cc	-!- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED	
Cc	POLY-N-ACETYLLACTOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I	
Cc	ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED	
Cc	WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS	
Cc	-!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL	
Cc	-1,4-N-ACETYL-D-GLUCOSAMINYLT-R = UDP + N-ACETYL-BETA-D-	
Cc	GLUCOSAMINYLT-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYLT-R.	
Cc	-!- PATHWAY: GLYCOSYLATION.	
Cc	-!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, GOLGI.	
Cc	-!- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2BNT).	
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration	
Cc	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
Cc	the European Bioinformatics Institute. There are no restrictions on its	
Cc	use by non-profit institutions as long as its content is in no way	
Cc	modified and this statement is not removed. Usage by and for commercial	
Cc	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
Cc	or send an email to license@isb-sib.ch).	
Dr	EMBL: U68182; AAB39621.1; -	
Dr	MGD: MG1:1100870; Gcnf2.	
Kw	Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;	
Ft	Gold1 stack; Glycoprotein.	
Ft	DOMAIN	
Ft	TRANSMEM	
Ft	1 6	
Ft	25	
Ft	CYTOPLASMIC (POTENTIAL).	
Ft	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)	
Ft	(POTENTIAL).	
Ft	LUMENAL, CATALYTIC (POTENTIAL).	
Ft	N-LINKED (GLCNAG. . .) (POTENTIAL).	
Ft	N-LINKED (GLCNAG. . .) (POTENTIAL).	
Ft	N-LINKED (GLCNAG. . .) (POTENTIAL).	
Ft	N-LINKED (GLCNAG. . .) (POTENTIAL).	
Ft	CARDHYD	
Ft	388	
Ft	SEQUENCE	
Ft	400 AA; 45507 MW; 193FE3882E3B9691 CRC64;	
Qy	Query Match	
Qy	Best Local Similarity	
Qy	Matches 156; Conservative	
Qy	5 ILYVEVSLSTSPFVRNRHYNKDEVFREVNCGSIYQDEPLEIKSLFIARRDLIDLEDD	
Qy	18 VCVVSYVLSEFGGDSYOXKLISDSVRLSOVCSEFFID-----GKSFRFLWNRKI-----	
Qy	65 VVAVTSPDCDIYOTLRGYAOKLVSKKEEKSFIAVSIVHKDAIVEERLIHAIFYNQENIYCI	
Qy	65 MIHEKPSCTIYYQSHITAPLSQEGSDPLAVVMVIHHNFDFEARFLRAIFMPQKITVCV	
Qy	125 HYDRKADPTFKVAMNNLAKCFSNIFLASKLEAVEYAHISLQADLCISDLKSSLOWKY	
Qy	125 HVDEKAATAEKRGAEVDVSCFPNVFLASKEPEVVYGISHSLOADLNLCKLSTSEVPWKY	
Qy	185 VINCGGDPLKSNFELVELSELKLNANMLETYPKPSKLERFTY--NHDLRRVPEYEVKL	
Qy	185 AINNGGDPLKTKNELVOYTLKGKGNLTPGVLPRAHAIGTRRVYRHSLKELSYV--	
Db	Score 730.5; DB 1; Length 400;	
Db	Pred. No. 28-46;	
Db	Mismatches 160; Indels 23; Gaps	

```

QY 244 PIRNISKREAPHNIOIEFGSAFVLSQAFVKYIFNNSIVDFEAMSKDYSDEHEFWAT 303
DB 243 -IRTTALKEPPPHNLITFGSAVALSRFANFVLDPRAVDLLQMSKDTFSPDEHEFWAT 301
QY 304 LIRVPIPEIISASQADVDLSQKTLVKWNYEGFFYSCGSHLRSCVIGAELEWML 363
DB 302 LNTPEVPSPPNVA-----WTGNLRAIKMSDMED-RHGCGSHYVHGICITGNGDLQWL 355
QY 364 IKDGHFWANKFDSKVDPIILKLAERLEQO 393
DB 356 INQSLSFANKFELNTYPLVECELELRER 385

RESULT 5
BIGB_HUMAN STANDARD: PRT; 400 AA.
AC 006430;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE N-ACETYLACTOSAMINIDE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE
DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)
DE (IGNT).
GN GCNT2 OR II.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93194065; PubMed=8449405;
RA Bierhuizen M.F.A., Mattel M.-G., Fukuda M.;
RT Expression of the developmental I antigen by a cloned human cDNA
RT encoding a member of a beta-1,6-N-acetylglucosaminyltransferase gene
RT family.
RL Genes Dev. 7:468-478(1993).
RN (2)
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=96078409; PubMed=7579796;
RA Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT Genomic organization of core 2 and I branching beta-1,6-N-
RT acetylglucosaminyltransferases. Implication for evolution of the
RT beta-1,6-N-acetylglucosaminyltransferase gene family.
RL Glycobiology 5:417-425(1995).
RN (3)
RP TISSUE SPECIFICITY.
RX MEDLINE=98070745; PubMed=9405606;
RA Sasaki K., Kurita-Miura K., Ujita M., Angata K., Nakagawa S.;
RA Sekine S., Nishi T., Fukuda M.;
RT Expression cloning of cDNA encoding a human beta-1,3-N-
RT acetylglucosaminyltransferase that is essential for poly-N-
RT acetylglucosamine synthesis.
RL Proc. Natl. Acad. Sci. U.S.A. 94:14294-14299(1997).
CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED
CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED
CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
CC CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL
CC -1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-
CC GLUCOSAMINYL-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, HIGHLY EXPRESSED IN PROSTATE AND
CC TO A LESSER EXTENT IN SMALL INTESTINE AND COLON. BARELY DETECTED
CC IN HEART, BRAIN, KIDNEY AND PANCREAS. NO EXPRESSION IN PLACENTA,
CC LUNG, LIVER, SKELETAL MUSCLE, SPLEEN, THYMUS, TESTIS, OVARY AND
CC PERIPHERAL BLOOD LEUKOCYTES. IN FETUS, HIGHLY EXPRESSED IN BRAIN
CC AND TO A LESSER EXTENT IN LUNG AND KIDNEY. BARELY DETECTED IN
CC LIVER.
CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION INCREASES DRASTICALLY DURING
CC DEVELOPMENT AND ONCOGENESIS.
CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).

```

```

CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L19659; AAA8177.1; -
DR EMBL; 219550; CAA79610.1; -
DR EMBL; L41607; AAA99832.1; -
DR EMBL; L41605; AAA99832.1; JOINED.
DR EMBL; L41606; AAA99832.1; JOINED.
DR PIR; A46297; A46297.
DR MIM; 600429; -
DR MIM; 110800; -
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Golgi stack; Glycoprotein.
FT TRANSMEM 7 25
FT DOKAIN 1 6
FT DOKAIN CYTOPLASMIC (POTENTIAL).
FT DOKAIN SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOKAIN (POTENTIAL)
FT DOKAIN LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 26 400
FT CARBOHYD 37 37
FT CARBOHYD 255 255
FT CARBOHYD 314 314
FT CARBOHYD 388 388
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 400 AA; 45854 MW; 1469628690A1F43B CRC64;

Query Match 32.4%; Score 711; DB 1; Length 400;
Best Local Similarity 43.7%; Pred. No. 5.3e-45;
Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;
QY 70 SDCDIQRLRGYAKIVSKKEKSPFIASVLYVHDAIVRELIHAIYNQNYCIHIDRK 129
DB 70 SSKKEYLTOSHTYAPLSKEADPEPLAIWYIHNFDTFARFLPAITMPCNTYVHDER 129
QY 130 APDFKVMNNLAKCFSNIFASKLEAVEYAHISRLQADLNCISDLKSSIQWKYVNLIC 189
DB 130 ATTEPKDAVEOLDSCFNAPFLASKMEPVYGGISRLQADLNCISDLKSAFVSMKXYVNTIC 189
QY 190 GODFPLKSNFELYSKTLKNGANNLEIYKPNKSLERTY-HHELRNRYPYVYKPLRTN 248
DB 190 GODFPLKSNFELYSKTLKNGANNLEIYKPNKSLERTY-HHELRNRYPYVYKPLRTN 248
QY 190 GODFPLKSNFELYSKTLKNGANNLEIYKPNKSLERTY-HHELRNRYPYVYKPLRTN 248
DB 190 GODFPLKSNFELYSKTLKNGANNLEIYKPNKSLERTY-HHELRNRYPYVYKPLRTN 248
QY 249 ISKEAPPHNIOIEFGSAFVLSQAFVKYIFNNSIVDFEAMSKDYSDEHEFWATLIRVP 308
DB 249 ISKEAPPHNIOIEFGSAFVLSQAFVKYIFNNSIVDFEAMSKDYSDEHEFWATLIRVP 308
QY 247 ALKPPPHNLITFGSAVALSRFANFVLDPRAVDLLQMSKDTFSPDEHEFWATLIRIP 306
DB 247 ALKPPPHNLITFGSAVALSRFANFVLDPRAVDLLQMSKDTFSPDEHEFWATLIRIP 306
QY 309 GIPGISISQADVDLSQKTLVKWNYEGFFYSCGSHLRSCVIGAELEWMLINDGH 368
DB 309 GIPGISISQADVDLSQKTLVKWNYEGFFYSCGSHLRSCVIGAELEWMLINDGH 368
QY 307 GVGSPMPNVA-----WTGNLRAIKMSDMED-RHGCGSHYVHGICITGNGDLQVNSPS 360
DB 307 GVGSPMPNVA-----WTGNLRAIKMSDMED-RHGCGSHYVHGICITGNGDLQVNSPS 360
QY 369 WFANKFDSKVDPIILKLAERLEQO 395
DB 369 WFANKFDSKVDPIILKLAERLEQO 395
QY 361 LEANKFELNTYPLVECELELRER 385
DB 361 LEANKFELNTYPLVECELELRER 385

RESULT 6
V$4_METJA STANDARD: PRT; 895 AA.
AC 058454;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN M1054 (EC 1.1.1.-).
GN M1054.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN (1)

```


RT "K-sam, an amplified gene in stomach cancer, is a member of the
RT heparin-binding growth factor receptor genes."
RL Proc. Natl. Acad. Sci. U.S.A. 87:5983-5987(1990).
RN [15]
RP SEQUENCE FROM N.A.
RX MEDLINE=92212948; PubMed=1313574;
RA Katon M., Hattori Y., Sasaki H., Tanaka M., Sugano K., Yazaki Y.,
RA Sugimura T., Terada M.;
RT "K-sam gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase."
RL Proc. Natl. Acad. Sci. U.S.A. 89:2960-2964(1992).
RN [16]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93016048; PubMed=1400433;
RA Dell K.R., Williams L.T.;
RT "A novel form of fibroblast growth factor receptor 2. Alternative
RT splicing of the third immunoglobulin-like domain confers ligand
RT binding specificity."
RL J. Biol. Chem. 267:21225-21229(1992).
RN [17]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX Steinberger D., Mueller U.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [18]
RP VARIANTS CS HIS-340; ARG-342; SER-342; TYR-342 AND CYS-354.
RX MEDLINE=95078932; PubMed=7987400;
RA Reardon W., Winter R.M., Rutland P., Pulleyn L.J., Jones B.M.,
RA Malcolm S.;
RT "Mutations in the fibroblast growth factor receptor 2 gene cause
RT Cronzon syndrome."
RL Nat. Genet. 8:98-103(1994).
RN [19]
RP VARIANTS CS CYS-328 AND CYS-347. AND VARIANT JWS GLY-344.
RX MEDLINE=95179174; PubMed=78741170;
RA Jabs E.W., Li X., Scott A.F., Meyers G., Chen W., Eccles M., Mao J.,
RA Charnas L.R., Jackson C.E., Jaye M.;
RT "Jackson-Weiss and Cronzon syndromes are allelic with mutations in
RT fibroblast growth factor receptor 2."
RL Nat. Genet. 8:275-279(1994).
RN [110]
RP VARIANTS CS
RX Oldridge M., Wilkie A.O.M., Slaney S.F., Poole M.D., Pulleyn L.J.,
RA Rutland P., Hockley A.D., Wake M.J.C., Goldin J.H., Winter R.M.,
RA Reardon W., Malcolm S.;
RT "Mutations in the third immunoglobulin domain of the fibroblast growth
RT factor receptor-2 gene in Cronzon syndrome."
RL Hum. Mol. Genet. 4:1077-1082(1995).
RN [111]
RP VARIANT PS ALA-321.
RX MEDLINE=95235551; PubMed=7719333;
RA Lajeunie E., Wei M.H., Bonaventure J., Munnich A., le Merrer M.,
RA Renier D.;
RT "FGFR2 mutations in Pfeiffer syndrome."
RL Nat. Genet. 9:108-108(1995).
RN [112]
RP VARIANTS AS TRP-252 AND ARG-253.
RX MEDLINE=95235562; PubMed=7719344;
RA Wilkie A.O.M., Slaney S.F., Oldridge M., Poole M.D., Ashworth G.J.,
RA Hockley A.D., Hayward R.D., David D.J., Pulleyn L.J., Rutland P.,
RA Malcolm S., Winter R.M., Reardon W.;
RT "Apert syndrome results from localized mutations of FGFR2 and is
RT allelic with Cronzon syndrome."
RL Nat. Genet. 9:165-172(1995).
RN [113]
RP VARIANTS PS PRO-341; ARG-342 AND TYR-342.
RX MEDLINE=95235563; PubMed=7719345;
RA Rutland P., Pulleyn L.J., Reardon W., Barisier M., Hayward R.,
RA Jones B., Malcolm S., Winter R.M., Oldridge M., Slaney S.F.,
RA Poole M.D., Wilkie A.O.M.;
RT "Identical mutations in the FGFR2 gene cause both Pfeiffer and
RT Cronzon syndrome phenotypes."
RL Nat. Genet. 9:173-176(1995).
RN [114]
RP VARIANTS CS/JWS/PS.
RX MEDLINE=96203627; PubMed=8644708;
RA Meyers G.A., Day D., Goldberg R., Daentl D.L., Przylepka K.A.,
RA Abrams L.J., Graham J.M. Jr., Feingold M., Moeschler J.B.,
RA Ramsley E., Scott A.F., Jabs E.W.;
RT "FGFR2 exon IIIa and IIIc mutations in Cronzon, Jackson-Weiss, and
RT Pfeiffer syndromes: evidence for missense changes, insertions, and a
RT deletion due to alternative RNA splicing."
RL Am. J. Hum. Genet. 58:491-498(1996).
RN [115]
RP VARIANTS CS CYS-105; GLU-338; CYS-351 AND ARG-384.
RX MEDLINE=97101656; PubMed=8946174;
RA Pulleyn L.J., Reardon W., Wilkes D., Rutland P., Jones B.M.,
RA Hayward R., Hall C.M., Brunton L., Chun N., Lammer E., Malcolm S.,
RA Winter R.M.;
RT "Spectrum of craniosynostosis phenotypes associated with novel
RT mutations at the fibroblast growth factor receptor 2 locus."
RL Eur. J. Hum. Genet. 4:283-291(1996).
RN [116]
RP VARIANTS CS ILE-331; ASP-336--ALA-337 DUPL AND TRP-356--THR-358 DEL.
RX MEDLINE=97114301; PubMed=8956050;
RA Steinberger D., Mulliken J.B., Mueller U.;
RT "Cronzon syndrome: previously unrecognized deletion, duplication, and
RT point mutation within FGFR2 gene."
RL Hum. Mutat. 8:386-390(1996).
RN [117]
RP VARIANT PS CYS-290.
RX MEDLINE=97295073; PubMed=9150725;
RA Tartaglia M., Valeri S., Velardi F., di Rocco C., Battaglia P.A.;
RT "Trp290Cys mutation in exon IIIa of the fibroblast growth factor
RT receptor 2 (FGFR2) gene is associated with Pfeiffer syndrome."
RL Hum. Genet. 99:602-606(1997).
RN [118]
RP VARIANT CS I-252. VARIANT AS F-252, AND VARIANT PS F-252--S-253.
RX MEDLINE=97156222; PubMed=9002682;
RA Oldridge M., Lunt P.W., Zackai E.H., McDonald-McGinn D.M., Muenke M.,
RA Moloney D.M., Twigg S.R.F., Heath J.K., Howard T.D., Hoganson G.,
RA Gagnon D.M., Jabs E.W., Wilkie A.O.M.;
RT "Genotype-phenotype correlation for nucleotide substitutions in the
RT Igit-Igit linker of FGFR2."
RL Hum. Mol. Genet. 6:137-143(1997).
RN [119]
RP VARIANTS CS GLU-292.
RX MEDLINE=97297373; PubMed=9152842;
RA Steinberger D., Collmann H., Schmalenberger B., Mueller U.;
RT "A novel mutation (A886G) in exon 5 of FGFR2 in members of a family
RT with Cronzon phenotype and plagiocephaly."
RL J. Med. Genet. 34:420-422(1997).
RN [120]
RP VARIANTS CS VAL-276; CYS-301 AND SER-314.
RX MEDLINE=98180879; PubMed=9521581;
RA Steinberger D., Vriend G., Mulliken J.B., Mueller U.;
RT "The mutations in FGFR2-associated craniosynostoses are clustered in
RT five structural elements of immunoglobulin-like domain III of the
RT receptor."
RL Hum. Genet. 102:145-150(1998).
RN [121]
RP VARIANTS AS TRP-252 AND ARG-253.
RX MEDLINE=98112406; PubMed=9452027;
RA Tsai F.-J., Hwu W.-L., Lin S.-P., Chang J.-G., Wang T.-R., Tsai C.-H.;
RT "Two common mutations 934C to G and 937C to G of fibroblast growth
RT factor receptor 2 (FGFR2) gene in Chinese patients with Apert
RT syndrome."
RL Hum. Mutat. Suppl. 1:S18-S19(1998).
RN [122]
RP VARIANT PS CYS-351.
RX MEDLINE=98358420; PubMed=9693549;
RA Mathijssen I.M., Vaandrager J.M., Hoogeboom A.J.,
RA Hesselink-Janssen A.L., Van den Ouweland A.M.M.;
RT "Pfeiffer's syndrome resulting from an S351C mutation in the

Query Match
Best Local Similarity 24.0%; Pred. No. 4.5;
Matches 78; Conservative 35; Mismatches 109; Indels 103; Gaps 18;

4.7%; Score 102.5; DB 1; Length 821;

FT VASPLIC 38 152 MISSING (IN SHORT ISOFORM).
FT VASPLIC 314 361 AAGVNTDEKEIVLIRNTEFEGAGECTGACNLSISIFHS
FT VASPLIC 314 361 AMTLVLP -> HSGINSSNVEALFNVTMDGECYKVS
FT VASPLIC 314 361 NYIGANOSAMTLVLPKQ (IN SHORT ISOFORM).
FT CONFLICT 169 169 C -> V (IN REF. 2).
FT CONFLICT 187 187 S -> P (IN REF. 2).
SQ SEQUENCE 821 AA; 91983 MW; FCD828ADD61F4414 CRC64;

144 CFSNFIASKL-----EAVEYAHISRLQADL--NCLSDL-----LKSSIQWKYVIN 187
DB 491 CFCGVMAEAVGIDKDKKEAVTVA-VKMLKDATEKDLSDVSEMEMKMKIGKKNIN 549
QY 188 L--CGGDFPLKSNFELSELKINGANMLETVPKPSKLERFTYHHLRVPYEVYKLP 244
DB 550 LIGACTOGPL---VYVEYASKGNLRYLAKRPQME-----YSIDINVPDECKTF- 600
QY 245 IRTNISKAPPHNIQIFVGSAYFV-----LSQAFYKIFNNSIVODEFAMSKDTYSP 296
DB 601 -----KDLVSCYOLARGMEYLSQKCIHRDLAARNVLTENNVMKXADFGALARDINI 654
QY 297 DEHFVATLIRVGIPELISRAQDVSDLSQKTRLYKNNVYEGFFYPSQSGHLSVCIYG 356
DB 655 DYKKKTNGRLP-----VKMMAPALDRYIT--HSDVMSFG 690
QY 357 -----AAELRWLIXDGMHANKFDSKVPILIKLAELKEOQRD-W 397
DB 691 VLAMEIFLIGSGPYGIVBELFKLKGH---RMDKPTN-----CTNE-LYMMWRDQW 740
QY 398 ITLSEK-LF-----MDRNLTTTS 415
DB 741 HAVPSQRPLFKQVEDRDLRLITLT 765

RESULT 11
BACB_BACLI STANDARD; PRT: 2607 AA.
ID BACB_BACLI STANDARD; PRT: 2607 AA.
AC 068007.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BACTIRACIN SYNTHETASE 2 (BA2) [INCLUDES: ATP-DEPENDENT LYSINE
ADENYLASE (LYSA) (LYSINE ACTIVASE); ATP-DEPENDENT D-ORNITHINE
ADENYLASE (D-ORNA) (D-ORNITHINE ACTIVASE); ORNITHINE RACEMASE
DE (EC 5.1.1.12)].
GN BACB.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;
RC MEDLINE=98089193; PubMed=9427658;
RA Konz D., Kleins A., Scherzendorfer K., Marahiel M.A.;
RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
RT 10716: molecular characterization of three multi-modular peptide
RT synthetases".
RT Chem. Biol. 4:927-937 (1997).
RL Chem. Biol. 4:927-937 (1997).
CC -1- FUNCTION: ACTIVATES TWO AMINO ACIDS AND INCORPORATE A D-ORNITHINE
CC FROM ITS SECOND ACTIVE SITE INTO BACTIRACIN.
CC -1- CATALYTIC ACTIVITY: L-ORNITHINE = D-ORNITHINE.
CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETHEINE
CC (POTENTIAL).
CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC ANTIBIOTIC BACTIRACIN.
CC -1- SUBUNIT: LARGE MULTISUBUNIT COMPLEX OF BA1, BA2 AND BA3.
CC -1- DOMAIN: CONSISTS OF TWO MODULES WITH A C-TERMINAL EPIMERIZATION
CC DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO THE PEPTIDE

CC PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR
CC SUBSTRATE ADENYLATION, THIOALATION, CONDENSATION (NOT FOR THE
CC INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND N
CC METHYLATION (OPTIONAL).
CC MISCELLANEOUS: BACTIRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC DODECAPETIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC ABUNDANT IS BACTIRACIN A, A BRANCHED CYCLIC DODECAPETIDE. IT
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-ILEU-D-
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC PHE-9, AND ASP-11).
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF007865; AAC06347.1; -
CC InterPro: IPR000255; -
CC InterPro: IPR001242; -
CC InterPro: IPR001242; -
CC Pfam: PF00501; AMP-binding; 2.
CC Pfam: PF00668; DUF4; 3.
CC Pfam: PF00550; PP-binding; 2.
CC PRINTS: PR00154; AMPBINDING.
CC PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
CC PROSITE: PS00455; AMP BINDING; 2.
CC PROSITE: PS00075; ACP DOMAIN; 2.
CC PROSITE: PS00075; ACP DOMAIN; 2.
CC Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
CC Multi-functional enzyme; Repeat.
CC REPEAT 535 1090
CC REPEAT 1547 2141
CC REPEAT 1021 1037
CC DOMAIN 1051 1051
CC BINDING 1051 1051
CC BINDING 2094 2094
CC SEQUENCE 2607 AA; 297474 MW; FF654FACB8BBA6F CRC64;

Query Match
Best Local Similarity 19.8%; Pred. No. 29;
Matches 85; Conservative 71; Mismatches 128; Indels 145; Gaps 25;

4.6%; Score 100.5; DB 1; Length 2607;
Best Local Similarity 19.8%; Pred. No. 29;
Matches 85; Conservative 71; Mismatches 128; Indels 145; Gaps 25;

QY 26 KDEYREYVNCSG--IYOEPLTIGKSLERRRDIIDLEDDDYVAMSDCDIYQTLRGYA 82
DB 606 EDYRYKMLEDSGADMYVIOEPRK---SKIDGRQLITAEQ-----TRFS 646
QY 83 QKLVSKERKSPFIYSLVH-----KDAIWER-LIHAIYONHNYCIHYDRKAP---- 131
DB 647 KENLPNKKADLAVLYITSSSGRPKGVMTTHNVH-----YDAFTIRIPLESH 698
QY 132 -----DTFKVANNLAKCFNSIFLASKLEAVEYAHISRLQADLCLSDLSKSI 180
DB 699 DTVLQVVSFSDAFSEEEYPIACSGRLVSRK-----SDLN-IDELVKTIG 745
QY 181 QMKVYINLCGGDFPLKSNFELSELKINGANMLETVPKPSKLERFTYHHLRVPYEVYKLP 244
DB 746 KYRTVLVSCS--PL-----LNLN-----DKQHLTFHPQKFISSGD 781
QY 237 --PYEVKLPRTNISKAPPHN---IQIFVGSAYVLSQAFYKIFNNSI-----VOD 285
DB 782 VLKREYV-----NIKGADVNSYSGPTEATVCAATYQLOSSADRK---KTSIPIGKPLSN 833
QY 286 FFAMSKDTYSPDEHFVATLIRVGIPELISRAQDVSD--LSQKTRLYKNNVYEGFFYPS 343

```

Db 834 YKVIADQYGRPOV-----GVPGELLIGEGVANGYLNHET-LTK-----AAFYVD 879
QY 344 CTGSHLRVCICGAEL-RWLKIDGHWPFANKPSKYDPIILKCLAELEOORDWITLPS 402
Db 880 ESEGER-----VYRTGDLARWLSDGNIEFLGRIDSQV---IGQYIELEIEE-----923
QY 403 EKLFMDRNL 411
Db 924 HRLMMDNI 932

RESULT 12
V001_VACCC STANDARD; PRT; 666 AA.
AC P21093;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE PROTEIN O1.
GN O1L.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_Taxid=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RL Virology 179:517-563(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M35027; AAA48053.1; -
DR PIR; D42510; DA2510.
SQ SEQUENCE 666 AA; 77577 MW; 46DF3299DDEB66EF CRC64;

Query Match 4.5%; Score 99.5; DB 1; Length 666;
Best Local Similarity 22.0%; Pred. No. 5.7; Mismatches 11; Indels 105; Gaps 18;
Matches 75; Conservative 50;

QY 8 VEYSLSTSPFRNRYTVHVKGEVREVCNCSGYEODEPLEICKSLERIR-----54
Db 292 VEVNISTYDF---RYROFADEFRIYIM---IKERQITMOSGRIKFRFRPMSLRSTIIR 345
QY 55 RDIDLED-----DDVY-----AMSDCIIYOTLRGTAQ 83
Db 346 KDTLSLEDILAHIDNARKNSKVSIEDVERLITSSFRNLPCVVRRTMLSDIDIKTKI--MWL 403
QY 84 KLVKKEKESFPIAYSLVHNKDAIWEVRLIAHYQNHICGHYDRK-----ADTFKVA 137
Db 404 KIV-KDMKSCALTLISAI---KGIVTDTINVLAK-----IHHNRNFKYLTSENNKEIA 455
QY 138 MNNLAKCFSNFIASKLEAVEYAHISRLQADLNC-----LSDLLKSSIQMKYVYNLCG 190
Db 456 VCNCSRCLSS-----LEFYRELKSVRCDLRFDGDLRLYLRLYALHKGKINQMLIG 505
QY 191 QDF--PLKSNFELVSELKKGANMLETVKRPNSKLEBFYTHHLLRVRPEYVYKLPRTN 248
Db 506 QRCWGPFLTEMLFENKKNKLN--NLMEYIKIS-----DMLYVGHSI-----EKLIFPIITDS 554

```

```

QY 249 ISKEAPPHNIQIFVGSAYFLVLSQAFVK-YIFNNSIYODEFA 288
Db 555 LSEKLSYDTMS-----VLNDQYAKIVIFENFTILEEYIA 587

RESULT 13
CEK3_CHICK STANDARD; PRT; 823 AA.
ID CEK3_CHICK
AC P18461;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TYROSINE KINASE RECEPTOR CEK3 PRECURSOR (EC 2.7.1.112).
GN CEK3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90332672; PubMed=2165604;
RA Pasquale E.B.;
RT "A distinctive family of embryonic protein-tyrosine kinase
RT receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:5812-5816(1990).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE DOMAINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M35196; AAA48655.1; -
DR PIR; B35963; B35963.
DR HSSP; P13362; IEG1.
DR InterPro; IPR000719; -
DR InterPro; IPR001245; -
DR InterPro; IPR003006; -
DR Pfam; PF00047; 1g; 3.
DR Pfam; PF00069; PKINASE; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW Transferrase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW Signal.
FT SIGNAL 1 23
FT CHAIN 24 823
FT DOMAIN 24 379
FT TRANSMEK 380 400
FT DOMAIN 401 823
FT DOMAIN 58 117
FT DOMAIN 174 240
FT DOMAIN 273 351
FT DOMAIN 135 145
FT DOMAIN 483 772
FT NP_BIND 489 497
FT BINDING 519 519
FT ACT_SITE 628 628
FT MOD_RES 659 659
FT DISULFID 65 110
FT DISULFID 181 233
FT DISULFID 280 344
FT CARBOHYD 86 86

```

```

FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 823 AA; 92299 MW; 42BP3CC4EA02FD43 CRC64;

Query Match 4.5%; Score 99.5; DB 1; Length 823;
Best Local Similarity 24.0%; Pred. No. 7.5; Indels 103; Gaps 18;
Matches 78; Conservative 34; Mismatches 110;

QY 144 CFSNFIASKL-----EAVEAHISRLQADL--NCISDL-----LKSSIQMKRYVIN 187
DB 493 CFGVVAEAVGIDKORPREAVTVA--VKMLKDATEKDLVSEMEMMKHKHNIIN 551
QY 188 L---CGDDFLKSNFELVSELKLNKGNMLETYKPPNSKJERTYHHLRRVPEYKLP 244
DB 552 LLAGCTDGGPL--YVIVYASKNREYLRARRPEME-----TSFDLRVPEQMTF- 602
QY 245 IRTNISKAPPHNIQIFVGSAYEV-----LSQAFVKYIFNNSIYQDFEAWSKDTYSP 296
DB 603 -----KDIVSCTQLARGHEVLASCKICHRDLAARVLTENNVMKIADFGIARDIINI 656
QY 297 DEHFMAFLIRPGIPEGISASQDVSLQSKTRYKKNYEGFFYPSCTSHLRVCYIG 356
DB 657 DYKKTNGRLP-----VKWMADEALFEDRYVT--HOSDWVSFG 652
QY 357 -----AAELRWLIRKGMFANKPESKYDPILIKCLAEKLEEQORD-W 397
DB 693 VLMEITFLGSPYGPGLPELLEFLKLEGH---RMDKPN-----CTNE-LVMWRHDCW 742
QY 398 ITLPSEK-LF-----MDRNIITTS 415
DB 743 QAVSORPTEFKQVEDLDRLITLT 767

RESULT 14
TEXT_CLOTE STANDARD; PRT; 1314 AA.
ID TEXT_CLOTE
AC P04958;
DR 13-AUG-1987 (Rel. 05, Created)
DR 13-AUG-1987 (Rel. 05, Last sequence update)
DR 15-JUL-1999 (Rel. 38, Last annotation update)
DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
OS Clostridium tetani.
OG Plasmid.
OC Bacterium; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87053814; PubMed=3536478;
RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA Weller U., Hudel W., Habermann E., Nhemann H.,
RT "Tetanus toxin: primary structure, expression in E. coli, and
RL homology with botulinum toxins."
RL EMBO J. 5:2495-2502(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-CN3911;
RX MEDLINE=87040747; PubMed=3774547;
RA Fairweather N.F., Lyness V.A.;
RT "The complete nucleotide sequence of tetanus toxin."
RL Nucleic Acids Res. 14:7809-7812(1986).
RN [3]
RP SEQUENCE OF 742-1314 FROM N.A.
RX MEDLINE=86085672; PubMed=3510187;
RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.,
RT "Cloning, nucleotide sequencing, and expression of tetanus toxin

```

```

RT fragment C in Escherichia coli."
RL J. Bacteriol. 165:21-27(1986).
RN [4]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=90201034; PubMed=2108021;
RA Kriegstein K., Henschen A., Weller U., Habermann E.,
RT "Arrangement of disulfide bridges and positions of sulfhydryl groups
in tetanus toxin."
RL Eur. J. Biochem. 186:39-45(1990).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE=92037649; PubMed=1935979;
RA Kriegstein K.G., Henschen A.H., Weller U., Habermann E.,
RT "Limited proteolysis of tetanus toxin. Relation to activity and
identification of cleavage sites."
RL Eur. J. Biochem. 202:41-51(1991).
RN [6]
RP IDENTIFICATION OF ZINC-PROTEASE.
RX MEDLINE=93010948; PubMed=1396558;
RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA Montecucco C.;
RT "Tetanus toxin is a zinc protease and its inhibition of
neurotransmitter release and protease activity depend on zinc."
RL EMBO J. 11:3577-3583(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
by proteolytic cleavage of synaptobrevin."
RL Nature 359:832-835(1992).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE=97475217; PubMed=9334741;
RA Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA Sax M.;
RT "Structure of the receptor binding fragment HC of tetanus
neurotoxin."
RL Nat. Struct. Biol. 4:788-792(1997).
CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-1-PHE-77
BOND OF SYNAPTOBREVIN-2.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-1-PHE-77 BOND IN
SYNAPTOBREVIN.
CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
AND ARE NON-TOXIC AFTER SEPARATION.
CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
GANGLIOSIDE RECEPTORS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
METALLOPROTEASES); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
SUBFAMILY.
CC -----
CC *THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04436; CAA28033.1; -
CC EMBL: M12739; AAA23282.1; -
CC EMBL: X06214; CAA29564.1; -
CC PIR: A25689; BTCLIN.
CC PDB: 1AF9; 29-APR-98.
CC PDB: 1A8D; 14-OCT-98.
CC MEROPS, M27.001; -.

```

```

DR InterPro: IPR000130; -
DR InterPro: IPR000395; -
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOKILYSIN.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
DR 3D-structure.
FT CHAIN 1 0
FT INIT MET 0
FT CHAIN 457 1314 TETANUS TOXIN LIGHT CHAIN.
FT METAL 232 232 TETANUS TOXIN HEAVY CHAIN.
FT ACT_SITE 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 236 236 BY SIMILARITY.
FT TRANSMEM 226 236 ZINC (CATALYTIC) (BY SIMILARITY).
FT TRANSMEM 669 689 POTENTIAL.
FT DISULFID 438 466 POTENTIAL.
FT DISULFID 1076 1092 INTERCHAIN.
SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EF81D CRC64;

Query Match 4.4%; Score 97.5; DB 1; Length 1314;
Best Local Similarity 20.0%; Pred. No. 20;
Matches 108; Conservative 78; Mismatches 154; Indels 201; Gaps 32;

QY 7 LVEYSISTPFPVNRRTYHVKDEYEVNCSGIYEOE-PLEICKSLERRRDIIDLED-D 64
DB 827 LMOYIRANSKFI--GITELK---KLESKINKVFSPIPFSYKNDCC---WVDNEEDID 877
QY 65 VYVMTS-----DCDIYOTLGRYACKVSKSEKSPFIAYSLVHKDAIWE---RLIH 113
DB 878 VILAKKSTIILNDINDIIDISGFS-----SVITPDAQLVGINGKAIH 923
QY 114 AIYNQHNITCIHYDRKAPDFFKVMNNLAKCFSNIF-----LASKLEAV---EYAHIS 163
DB 924 LVNNESESEYIVH---KAMD---LEYNDMFNNFTVSWFLRVPRVSAHSLEOYGTNEYSIIS 977
QY 164 RLQADLNCIS-----DLKLSIQMKY-----VINLGGDFPLKSNFELVSEL----- 205
DB 978 SMK--HSISIGSGMSVSLKGNLITLKDAGEVNOITFRDLPDKFNAYLANKVFTTI 1035
QY 206 --KLLGANK-----LETVRPNS--KLEFTHHE----- 232
DB 1036 TNDRLSSANLYINGVLMSAEITGLGAIREDNNITLKLRCNNNQYVSIDKRLFCAL 1095
QY 233 -----LR-----RVEYEVKLPDIRINKEAPHNI----- 258
DB 1096 NPKIEKLYTSYSTFLRDPENGNPLRYDTEYLILPVASS--SKDVLAKNITDYMYLITNAP 1154
QY 259 -----QIFVGSAYEVLQSAFYKIFNNSIYODFFAMSKD-----TYSDEHEFWA-- 302
DB 1155 SYTNKLTNIYRRLNGLKFIIRKRYPNNEI--DSFVKSGDFIKLVSYNNNEHIYGPYK 1212
QY 303 -----TLIRV-----GIRGEISRSQDVSDLOSRIYKWNMYE-----GFFYPS 343
DB 1213 DGAENLNDRLILVGNVAPGIPLYKMEAVKLDL--KITYVOLAKITDYKMSASGL----- 1266
QY 344 CTGSHRSYVCIYGAELRMLIKDGHWFANKFDSKVDPIILIKLAELKEQORDWITLPS 403
DB 1267 -VGTHNQOI---GNDFNRDILIASNMVYFNHLKDKI---LGC-----DMYVTPYD 1308
QY 404 K 404
DB 1309 E 1309

```

```

RESULT 15
YCF2_PINTH
ID YCF2_PINTH STANDARD: PRT: 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 244.6 KDA PROTEIN (ORF 2054).

```

```

GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Coniferales; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakaugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugiura M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D17510; BAA04460.1; -
DR InterPro: IPR001939; -
DR Pfam: PF00004; AAA; 1;
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;

```

```

Query Match 4.4%; Score 97.5; DB 1; Length 2054;
Best Local Similarity 21.9%; Pred. No. 36;
Matches 70; Conservative 55; Mismatches 86; Indels 109; Gaps 20;

QY 4 DIVYEVSLST--SPFVR-NRRTYHVKDE-----VREYVN--CSGIYEOEPLTIGK 48
DB 871 DLVR-SFVLRLNLPFRKRYLSIEISITPLTKQIVNFEKNFCOPFKRSDSENN 929
QY 49 SLEIRRRDILEDDDVYAMTSCDIYOTLGRYAKVYS-----KEKSPFIAYSLVHK 103
DB 930 FDQCFKR-----GFSSNGLIQT-RSYODDLISEMFSKNEIEFPTI-----Q 971
QY 104 DAIVERTLHAIVQHNITCIHYDRKAPDFFKVMNNLAKCFSNIFLASKLEAVEYAHIS 163
DB 972 DMEVTECKKNIVM-----DIDGRS-----TLSNKSKEQNIYRISQIDSI---FS 1015
QY 164 RLQADLNCISDLKSSSIOW-----KIVNLGGDFPLKSNFELVSELKLNKAN----- 212
DB 1016 KW-----DLFTYKMFFTSAMKRYENML-----LDTLSEI-LHGSNFPVS 1057
QY 213 MLETKPPNSKLEFRTYHHLRVRVYEVKLPDIRINISKEAPHNIQIFVGSAYEVLQSA 272
DB 1058 ILQNK--HNILKR--NIMELSHPLMERIQQCKLTNIN----- 1094
QY 273 FVKYIFNNSIYODFFAMSKD 292
DB 1095 --KEFFPSNNKDFEPYCKD 1112

```

Search completed: August 2, 2001, 14:50:20
Job time: 300 sec

Fri Aug 3 09:38:18 2001

us-09-645-192-2_copy_39_453.rsp

GenCore version 4.5
Copyright (c) 1993 - 2000 Comphen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2001, 14:45:50 ; Search time 25.93 Seconds
(without alignments)
1219.145 Million cell updates/sec

Title: US-09-645-192-2_COPY_39_453

Perfect score: 2194
Sequence: 1 POKDIYLVESLSTSPFVRN.....DWITLPSEKLFMDRLTTTS 415

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR-68:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	849.5	38.7	428	2 A46293	beta-1,3-galactosyl
2	711	32.4	400	2 A46297	beta-1,6-N-acetylgl
3	331.5	14.7	445	2 T21262	hypothetical prote
4	313.5	14.3	425	2 T20297	hypothetical prote
5	296	13.3	401	2 T24929	hypothetical prote
6	291	13.3	454	2 T21261	hypothetical prote
7	288.5	13.1	472	2 T22188	hypothetical prote
8	265	12.1	402	2 T24930	hypothetical prote
9	245	11.2	459	2 T24742	hypothetical prote
10	226	10.3	470	2 T25382	hypothetical prote
11	218.5	10.0	470	2 T32137	hypothetical prote
12	218	9.9	489	2 T24013	hypothetical prote
13	213.5	9.7	486	2 T21566	hypothetical prote
14	196.5	9.0	472	2 T33364	hypothetical prote
15	196	8.9	753	2 T24745	hypothetical prote
16	194.5	8.9	448	2 DB5042	probable glycosyla
17	188	8.6	478	2 T27714	hypothetical prote
18	187	8.5	447	2 T00906	hypothetical prote
19	182.5	8.3	513	2 T20068	hypothetical prote
20	178	8.1	367	2 T28892	hypothetical prote
21	178	8.1	467	2 T21408	hypothetical prote
22	175.5	8.0	406	2 F96571	hypothetical prote
23	168	7.7	395	2 B96735	unknown protein F2
24	166.5	7.6	434	2 T51450	hypothetical prote
25	132	6.0	630	2 T02524	probable RING zinc
26	131	6.0	384	2 T08940	hypothetical prote
27	123.5	5.6	371	2 T21819	hypothetical prote
28	116	5.3	895	2 E64431	UDPglucose 6-dehyd
29	108.5	4.9	661	2 S50734	WH3 protein - yea

30	108	4.9	702	2 E69498	hypothetical prote
31	106.5	4.9	1347	2 T41321	rib domain and Ank
32	104	4.7	2485	2 T48628	hypothetical prote
33	104	4.7	2628	2 H71621	serine/threonine-s
34	104	4.7	588	2 S59413	probable membrane
35	103.5	4.7	415	2 H82877	hypothetical prote
36	102.5	4.7	415	2 T65223	heparin-binding fi
37	102.5	4.7	707	2 A54846	fibroblast growth
38	102.5	4.7	707	2 A38429	keratinocyte growth
39	102.5	4.7	769	2 S16236	fibroblast growth
40	102.5	4.7	820	2 S17295	fibroblast growth
41	102.5	4.7	821	2 TVMSBK	fibroblast growth
42	102.5	4.7	821	2 TVHUF2	fibroblast growth
43	102.5	4.7	822	2 A45081	fibroblast growth
44	102.5	4.7	822	2 B54846	fibroblast growth
45	100.5	4.6	437	2 T23652	alpha-1,3-mannosyl

ALIGNMENTS

RESULT 1
A46293
beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A46293
R:Biochem. Acad. Sci. U.S.A. 89, 9326-9330, 1992
A:Title: Expression cloning of a cDNA encoding UDP-glucNAc:Gal beta 1-3-GalNAc-R (GlcN
A:Reference number: A46293; M01D:93028457
A:Accession: A46293
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-428 <BIB>
A:Cross-references: GB:M97347; NID:9183440; PIDN:AAA35919.1; PID:9183441
A:Experimental source: HL-60 cells
A:Note: sequence extracted from NCBI backbone (NCBIN:115900, NCBIPI:115901)
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	38.7%	Score	849.5;	DB 2;	Length	428;
Best Local Similarity	43.0%	Pred. No.	1.1e-56;			
Matches	161;	Conservative	70;	Mismatches	129;	Indels 41; Gaps 11;
QY	1	POKDIYLV- EYSLSTSPFVRNRYTHVKE- -VRY- - - - -	EYVNSGTYEQPLEI	46		
DB	14	PTKYEFKVLVSLITFSVLR- - - IHOKPEFVSRYHLEAGNPSSDINCTKYLGQGVNEI	70			
QY	47	GK- - - - - SLEIRRDIDLEDVAVMTSDCDIYOTLKGYAKTVSKSEKSPFIASLV	100			
DB	71	QVKKLELLIVKFKRP- - - RPTPDYINMTSDCSFIRKKRYIEP-LSKEAEPIASLV	128			
QY	101	VHKDAIWERLLIATYNQNIYCIHYDKRAPDTFKYAMNNLACSENFIFASLEVEYA	160			
DB	129	VHNIEMIDRLALAIYMPQNFYCVHDTKSEDSYLAVMGIASCFNVAASLESVVA	188			
QY	161	HISRLDNLCLSDILKSSIQMKYVIVNLGQDPLKSNFVSELKLLGANNLETVKPP	220			
DB	189	SWSVQADLNCMDYLAAMSANKYILNLCMDPIKTNLEIVAKLLMGNNLETERMP	248			
QY	221	NSKLEPFTYHHELRARYEVY- - - KLPIRINISKEADPHNIQIVGSATFYVLSQAFKYIF	278			
DB	249	SHKEER- - - - - KRRYEVNGL- - - TNGTVMKLPLETPLEPSGSAFYFVSREYGYVL	300			
QY	279	NSNIVQPFAMSKDYSDEHFNATLIRVGPIGETISRSAC- - - DVSQDSKTRFLVKNVYE	337			
DB	301	QNEKIQKLMEADOTYSDEITMATIORIDEVQSLASIKYDLSMQAARVKNQYFE	360			
QY	338	GFE- - - - - YPSCGSHLSVCYIGAAELRWLIKDGHWFAKPFQSKYDPIIKCLAEKLE	392			
DB	361	GDVKGAPYPPCDGVHVSVCIFGAGDINMMLRKHHLFANKFVDVDLFAIIOCLDEHLRH	420			


```
Db 67 ILAQILKSGENACQAEFGKIFGFOEPTSGEELERFLAYGMLVHGDFVQISLLSATY 126
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 117 NQNIYCIHIDRKADPTFKYAMNNLAKCFNSI--FIASKLEAVEYAHISLQADLNCSD 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 QPONQFLAVDGNSSVFEIGLVRLSRGYNIGYFTTDEIRMGCELTLSV--FQCVDY 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 175 LKSSIQMKYVINLCGODFPLKSNFELVSELKLCNGANMLETVKPPNSKLEFTHHELR 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 LAKIPSMWKTFQYILSGVDAPLKSLEMLIRLKLKNG-----SFAEIT- 225
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 235 RVPREYVKLPRTNISKADPPHNIQIFVGSAYFVLSQAFYKIFPNSIYODFFAMSKDY 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 -LPEFPLRNKRKRPWSSPLPLYKTSL--SATF--SRKSANFVNSERKYLEQIDFLRGTY 279
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 295 SPDEFWATLIRVPGI-----PGETSRSAQDVSDL-QSKTRLVK-----NYE 337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 CADSLWATINGNKKVILKFSLEPMGCFDAKAWIKHKYRTRRGKLGKGENOKINDNGYV 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 338 GFERYP-----SCTGSHLRNSVCYGAELRWLIKDGHWTFANKFDSKVDPILIKLAEL 380
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 SRYQOYVRAAPVKCKGYRSLCVFGYDLPNLINRHELVANHKLIFSYPAPAEWCLVENS 399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 391 EEQ 393
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 ROK 402
```

RESULT 5

```
T24929
hypothetical protein T15D6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24929
R: Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Accession: T24929
A:Reference number: 219956
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-401 <WIL>
A:Cross-references: EMBL:Z83125; PIDN:CA805620.1; GSPDB:GN00019; CESP:T15D6.2
A:Experimental source: clone T15D6
C:Genetics:
A:Gene: CESP:T15D6.2
A:Map position: 1
A:Introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3
```

```
Query Match 13.5%; Score 296; DB 2; Length 401;
Best Local Similarity 23.4%; Pred. No. 7.2e-15;
Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;
```

```
Qy 23 THVNDVAYEVNCSGYEQEPLKSGLEIRRDIDLEDDVYAMTSDCIIYOTLNGYA 82
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 TDLDDL--ELNGTNIL--QGLKNEKOLELINTKMT--EKLKNSIDRCHTVKSMRFEN 86
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 83 QLVSKSEKSPPIAYSLVYHKDAIYERLIHAIYNQNIYCIHIDRKADPTFKYAMNNLA 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 EYPLSEEEARFPLSYGLLYKELSOVLPMLSIYOPONECYIAGENSATFLLLELS 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 143 KCFSNIFASKLEAVEYAHISLQADLNCSDLKSSIQMKYVINLCGODFPLKSNFELY 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 DCFPN----KRPITWGSYETINSYDCLFLSHLKNMKYFQYLSGVDPILKTNLEKV 201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 203 SELKLCNGANMLETVKPPNSKLEFTHHELRVPEYVKLPRTNISKADPPHNIQIFV 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 202 RLKLSGANGANVEIKVEYERRL-----LQNETESPLPLFX 237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 263 GSAYFVLSQAFYKIFPNSIYODFFAMSKDYSPDEHFWATLIRVPGI--IPGETSRSAQ 319
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 SSLSLIPRKANYLASSSIPOOLLLEFLNMTVADEGFWGLGNKGLDVPVGSIN---- 293
```

```
Qy 320 DVSDLOSKTRLVKMYRYEGFYPSCGTGSHLRNSVCYGAELRWLIKDGHWTFANKFDSKV 379
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 -----FEEHQWFESGCHNMDQSCVFGIGVSNLLQAKALVAHKLYLTSE 340
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 380 PILIKCL 386
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 PEAYFCL 347
```

RESULT 6

```
T21261
hypothetical protein F22D6.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21261
R: Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: 219397
A:Accession: T21261
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-454 <WIL>
A:Cross-references: EMBL:Z71262; PIDN:CA95816.1; GSPDB:GN00019; CESP:F22D6.12
A:Experimental source: clone F22D6
C:Genetics:
A:Gene: CESP:F22D6.12
A:Map position: 1
A:Introns: 62/3; 92/2; 190/3; 266/3; 291/3; 412/3
```

```
Query Match 13.3%; Score 291; DB 2; Length 454;
Best Local Similarity 24.1%; Pred. No. 2e-14;
Matches 99; Conservative 73; Mismatches 174; Indels 64; Gaps 14;
```

```
Qy 17 EVARNRYTHVXDEV---RYEVNCSGYEQEPLKSGLEIRRDIDLEDDVYAMTSDC 73
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 40 FLKSICTTASDSYLLDNMEINCSNII--KGYKTNEKIDIMHDI--EQLFSCITNCCQ 94
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 74 IYOTLNGYAKLVSKSEKSPPIAYSLVYHKDA-----INVERLIHAI 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 TLKTLFRFNTNPMASAEKHFPLSYGLMYKDDLPQYPARMFILKHELEINLVFLLSI 154
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 116 YQNIYCIHIDRKADPTFKYAMNNLAKCFNSIYFIASKLEAVEYAHISLQADLNCSD 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 YHPQNECYIAGENSAPIFQNLREYSTCFSNVF--MKRPPISWGSHEIIDSVDLEFL 213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 176 LKSSIQMKYVINLCGODFPLKSNFELVSELKLCNGANMLETVKPPNSKLEFTHHELR 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 SHLETDRIYQYLSGVDPILKTNLEVYQILKNGTSNVEIT--NYQARLIGKNE--- 267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 236 VPEYVKLPRTNISKADPPHNIQIFVGSAYFVLSQAFYKIFPNSIYODFFAMSKDY 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 -----NESP---LPLFKSSLSALIPRKANQULASSITARKLLEFLNTEI 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 296 PDEHFWATLIRVPGI--IPGETSRSAQDVSDCQ--FLYVKNY-----EGFFYSCG 346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 ADEGFWGLGNKDDQRISSINSKMDMEYRDQNNIINFPTDQMSYIISRDQIMDELCK 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 347 SHLR-SVCYIAGAELRWLIKDGHWTFANKFDSKVDPILIKLAELDEQR 395
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 NYMKDSCVFGIGVDPILKTNLSKALVAHKFYLKSEPEAYFCL--LKEHNR 416
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 7
T22188
hypothetical protein f44F4.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22188
R: Coles, L.
submitted to the EMBL Data Library, September 1994
A:Reference number: 219528
A:Accession: T22188
```

A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-472 <WIL>
 A:Cross-references: EMBL:Z37092; PIDN:CA85457.1; GSPDB:GN00020; CESP:F44F4.6
 A:Experimental source: clone F44F4
 C:Genetics:
 A:Gene: CESP:F44F4.6
 A:Map position: 2
 A:Introns: 57/2; 105/2; 185/3; 261/3; 319/1; 370/1; 412/3

Query Match 13.1%; Score 288.5; DB 2; Length 472;
 Best Local Similarity 22.0%; Pred. No. 3.3e-14;
 Matches 91; Conservative 78; Mismatches 175; Indels 69; Gaps 13;

QY 40 EDEPLEIGKSLF--IRRDI-----IDLEDDVAMTSDCDIYQTLRGYAKLVSK 88
 Db 63 ENPLECNSVINGENRRKISRARQMDHFDWEHEIFNSRWVCSIDKXFNTRIPSSR 122
 QY 89 EESFPIAASLYHKDAIWERLIHAIYQNHNYCIHYDRKAPDFKVMNNLACFSN- 147
 Db 123 EEAEYPLAAGLVVYKTIYQVLTQMSLFYQPHFCITVDQSPNEKSVYQALPSCFFPM 182
 QY 148 -IFIAKLEAVEYAHISRLQADLNCISDLKSSIQKRYVINLCGDFPLKSNFELVSEK 206
 Db 183 HAFIG--EPSQMSGILKNVYTCFNLKSKQKWKYQYLSGTDLPRTNLEWVRIFK 239
 QY 207 KLNGAMLEVKPPNSKLEFTHHLRVRPYVYVLPRTINISKAPPHNIQIFGSAV 266
 Db 240 ALNGS-----MNTDVSTF---FVDRP-----KNNEGVLP--HPVYKSSMS 275
 QY 267 FVLSQAFVKYIFNNSIVODEFAMSKDYSPDEHFMAFLIRVPD--IPEISRSADVSD 323
 Db 276 VVVPREBADVLISSPRQKTLKLTWIPDESFMTVJGSPALLPVPST--RVMDILM 333
 QY 324 LQSKTRL-----VKNNYIEG---FYPSCTGSHLRSCVITGAAELRMLKDGHN 369
 Db 334 LRKFKLRPEYENTVNSIGSYGRQVWGMOKECFKQKDFSCVGVGVDIEBIMRPEL 393
 QY 370 FANKFDSKVPDILIKLAETLEKQQ-----RDMITLPSKLFEMRNL 412
 Db 394 VAHKLTYEFQPAFMCLKEVRRKRSLSPAHLFSARSYSQMPYVELYQKKAIR 446

RESULT 8
 T24930
 hypothetical protein T15D6.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24930
 R: Dobson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19956
 A:Accession: T24930
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-402 <WIL>
 A:Cross-references: EMBL:Z83125; PIDN:CA805621.1; GSPDB:GN00019; CESP:T15D6.3
 A:Experimental source: clone T15D6
 C:Genetics:
 A:Gene: CESP:T15D6.3
 A:Map position: 1
 A:Introns: 55/3; 85/2; 121/3; 161/3; 237/3; 262/3

Query Match 12.1%; Score 265; DB 2; Length 402;
 Best Local Similarity 24.7%; Pred. No. 1.6e-12;
 Matches 93; Conservative 64; Mismatches 157; Indels 62; Gaps 13;

QY 23 THVDEVEYVNCGIVOEPLIGKSLERIRRDIDLEDDVAMTSDCDIYQTLRGYA 82
 Db 44 TDLDDL--QINCINILGFKNE--NTELELINTKMT--EKMMNSNDRCOTLISMRFN 96

QY 83 QKLVSKKEKSPPLAYSILVHKDAIWERLIHAIYQNHNYCIHYDRKAPDFKVMNNL 142
 Db 97 KVPISSEEARPLPSFGILVYKELSOVLFLLSSITQPNQNEICIVGENSAPAFILLKEIA 156
 QY 143 KCFENIFIAKLEAVEYAHISRLQADLNCISDLKSSIQKRYVINLCGDFPLKSNFELV 202
 Db 157 NCFPN-----KRPPIKNGSPEIINSYVGCLEFSLHLSKMDKRYOYLSGVDIPLKTNEMV 211
 QY 203 SELKLGAMLEVKPPNSKLEFTHHLRVRPYVYVLPRTINISKAPPHNIQIFV 262
 Db 212 RLKRLNG-----TVNIGIS-----TYEDRL-----LNGKNTESP--LPLFK 247
 QY 263 GSATFVLSQAFVKYIFNNSIVODEFAMSKDYSPDEHFMAFLIRVPDIPGHSR 316
 Db 248 SLSLILPRKANYLSSSSVQQLLEFLRTTWADSEGFMTGLGNKDLFVNPSPFNDR 307
 QY 317 SAQVSDLOSKTRLVKM-----NYEGFYPSCGSHLRSCVITGAAELRMLKDGHN 370
 Db 308 LTNGMGNVYSRHQL--WVESECHNYMKD-----RS-CVFGIGDVNLMKSRALV 353
 QY 371 ANKFDKVPDILIKCL 386
 Db 354 AHKLTYESEPEAFCL 369

RESULT 9
 T24742
 hypothetical protein T09E11.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T24742
 R:McLay, K.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19930
 A:Accession: T24742
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-459 <WIL>
 A:Cross-references: EMBL:Z81147; PIDN:CA803336.1; GSPDB:GN00019; CESP:T09E11.9
 A:Experimental source: clone T09E11
 C:Genetics:
 A:Gene: CESP:T09E11.9
 A:Map position: 1
 A:Introns: 39/1; 78/3; 204/1; 291/1; 338/3; 371/3; 417/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 11.2%; Score 245; DB 2; Length 459;
 Best Local Similarity 22.8%; Pred. No. 6.2e-11;
 Matches 106; Conservative 74; Mismatches 202; Indels 82; Gaps 17;

QY 5 IYLVKSLSTSPFVNRNRTHVDEVREYVNCGIY--EOEPLIGKSLERIRRDIDLED 62
 Db 25 VYLMFS--SKSLFGRN-----EONSVALTNLENENQRAIINELRPRFAKIES 72
 QY 63 -----DDVAMTSDCDIYQTLRGYAKLVSKESFSP----- 94
 Db 73 KIEHFYSKYRRRSEETAHVDCGRILSGCKDYQTVSG--ENRIPVEMPNPDMSCSAVMD 130
 QY 95 -----IAYLVYHKDAIWERLIHAIYQNHNYCIHYDRKAPDFKVMNNL 139
 Db 131 RIIPHDILIRPKNGVAEFRIYKDEYKOVQMSYHPSFCALDKRAIDKRAFRKQMR 190
 QY 140 NLAQFSNIFIAKLEAVEYAHISRLQADLNCISDLKSSIQKRYVINLCGDFPLKSNF 199
 Db 191 AMAACIPVLLLPDEPIDSNHNVNLAHLNCLRALINKP--GMNYAMLLQNHDLTLKSYV 249
 QY 200 ELVSEKLGAMLEVKPPNSKLEFTHHLRVRPYVYVLPRTINISKAPPHNIQ 259
 Db 250 ELEQVYEWLGANDVELL--PEAQRLDEENFKWD---PSLAKFPDESQVDETILNEKIK 304
 QY 260 IFVGSAYFVLSQAFVKYIFNNSIVODEF--AMSKDYSPDEHFMAFL--IRVPGIGELISR 316

Db 305 FSKGVQSGSKRAAVDMTRKYNLSITYIDMNGRWGVDEMLISSIQISAFGLMPGHF-- 362
 QY 317 SADDVSDLSKTRLVKMYEYGFYSCGSHLR-SVCITYGAELRWIKDGHWFANKPD 375
 Db 363 TDGCLRGKPFPMQWNEFDESY---CASKYRHNVCILGIELYSVASFPILMNKML 419
 QY 376 SKVDPILIKCLAELEQ---ORDWITLPSEKLFMDRLTTTS 415
 Db 420 PPDNSIIECTAEILYNRTFMQNDH---PLEEYV-KNNVTVS 459

RESULT 10
 T2382
 hypothetical protein T27F6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T2382
 R:Dobson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z20026
 A:Accession: T2382
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-454 <MIL>
 A:Cross-references: EMBL:282060; PIDN:CAB04881.1; GSPDB:GN00019; CESP:T27F6.1
 A:Experimental source: clone T27F6
 C:Genetics:
 A:Gene: CESP:T27F6.1
 A:Map position: 1
 A:Introns: 80/3; 191/1; 278/1; 325/3; 365/2; 407/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 10.3%; Score 226; DB 2; Length 454;
 Best Local Similarity 24.4%; Pred. No. 1.7e-09;
 Matches 80; Conservative 56; Mismatches 158; Indels 34; Gaps 12;

QY 84 KLVKEEKSPF---IAYSLVHKDAIMVERLHAIVNQHNYICINHYRKAPOTFKYAMN 139
 Db 110 RILSRDYLRLNGVAFARVYVMDYELIEKHEMSYHPQNSCFALDKAAKEFKRMQ 169
 QY 140 NLAKFSNIF-----ASKLEAVEYAHISRLQDLNCLSDLSKSQWKYVNLG 190
 Db 170 AMASCLPNVLLPGRFKNPHIDLSVDSHGHTNL-AHYNCLRALINKP-GWNPAILLON 227
 QY 191 ODFPKSFEELVSELKILNGANMLETVKPPNSKLERFTYHHELRVPEYVKLPRTNIS 250
 Db 228 HDLTKSYELEKIFNMIMGANDY-AIRPELGRDK-----KHFKMDMSLKLFENES 279
 QY 251 KEAP---PHNIQIVGSAFYVLSQAFVKYIFNNSIVODFF-AMSKDTYSPDEHFMAT--L 304
 Db 280 EIDVYIINTLTILKFAKGAOQSLSRAAVDMTRVDTLTFFIDMNGHGVDEQTOAFQI 339
 QY 305 IRVGIPEIS-RSAQDVSDLSKTRLVKMYEYGFYSCGSHLR-SVCITYGAELRW 362
 Db 340 SDFLGMPEHFDCKIKKGIITEGITRFQWTHGD---QSKAKSKSHGICIMGIEHLSM 396
 QY 363 LIKDGHWFAKFSKVDPIILIKCLAEKL 390
 Db 397 MAKSEHLMFNKVLPLFDYSIIECTAEILL 424

RESULT 11
 T2137
 hypothetical protein R07C3.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T2137
 R:Lamar, B.; Kramer, J.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid R07C3.
 A:Reference number: Z21125
 A:Accession: T2137

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-470 <LAM>
 A:Cross-references: EMBL:AF016686; PIDN:AAB66231.1; GSPDB:GN00020; CESP:R07C3.3
 A:Experimental source: strain Bristol N2; clone R07C3
 C:Genetics:
 A:Gene: CESP:R07C3.3
 A:Map position: 2
 A:Introns: 18/1; 78/3; 189/1; 276/1; 319/3; 359/2; 401/3; 441/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 10.0%; Score 218.5; DB 2; Length 470;
 Best Local Similarity 21.2%; Pred. No. 6.5e-09;
 Matches 90; Conservative 85; Mismatches 188; Indels 61; Gaps 17;

QY 22 YTHVKDEVR-YEVNC-SGIYEQEP-----LEIKSLEIRRRDI--DLEDDVYA 67
 Db 39 YKRIEALRTVDRCSEEGGFYCKRPETQHVDCGRVLVDKIQLRFAFYQLSLGSGSNIP 98
 QY 68 MTSDDDIYQTLRGVAKIYKSEKSPF-----AYSLVHKDAIMVERLHAIVNQHNY 122
 Db 99 LIENSFLNLTGSAIKSRITPTAOKFPLKNGTAFARIVPADYEIEKQVQASHYPQNF 158
 QY 123 CIHYDKAPDIFKVMNMLAKCFNIFLASKLEAVEYAHISRLQDLNCLSDLSKSQW 182
 Db 159 CRAIDANSSAEQRMKKKLEQCLPNNVVLPTYESYDSGHNINLAHYNCMKK-LESIRGW 217
 QY 183 KYVINLGGDFPLKSNFELVSELKILNGANMLETVKPPNSKLERFTYHHELRVPEYVK 242
 Db 218 GYLMLONHNDVITKSYELDRIFELLGVN-----DVEKSRITPRKKHKLMDLKS 269
 QY 243 LPRTNISKAPPHNIQIVGSAFYVLSQAFVKYIFNNSIVODFF-AMSKDTYSPDEHFM 301
 Db 270 LKLFNRDKNNS-DYELIISGVSHSLSRAAYKMLEVDLSTFIDQNMNQDYGADQFI 328
 QY 302 AFLIRVPI--PGEISRA-QVSDLSKTRLVKMYEYGFYSCGSHLRASVCITGA 357
 Db 329 STFQMSPDLMKMGHFNEDICINDTAIVTISRLARWPYILDRAAHCACTVHRHDCITGI 388
 QY 358 AELRWLIKDGHWFAKFSKVDPIIL-----IKCLAELEQORDMWITLPSEKLFM 407
 Db 389 EDPRAI-----SKF-----PILNMLPAFDYSIIECTAEILL--HNRYYLGQVQDKI-- 431
 QY 408 DKNL 411
 Db 432 EKNL 435

RESULT 12
 T24013
 hypothetical protein R07B7.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24013
 R:Harits, B.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19830
 A:Accession: T24013
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-489 <MIL>
 A:Cross-references: EMBL:275955; PIDN:CAB00115.1; GSPDB:GN00023; CESP:R07B7.6
 A:Experimental source: clone R07B7
 C:Genetics:
 A:Gene: CESP:R07B7.6
 A:Map position: 5
 A:Introns: 42/3; 80/2; 235/3; 287/1; 372/2; 416/1

Query Match 9.9%; Score 218; DB 2; Length 489;
 Best Local Similarity 22.0%; Pred. No. 7.5e-09;
 Matches 97; Conservative 74; Mismatches 197; Indels 72; Gaps 18;

hypothetical protein T09E11.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24745

R:McClay, K.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19930

A:Accession: T24745

A:status: preliminary; translated from GB/EMBL/DBJ

A:molecule type: DNA

A:Residues: 1-753 <WIL>

A:Cross-references: EMBL:Z81147; PIDN:CAM03539.1; GSPDB:GN00019; CESP:T09E11.6

A:Experimental source: clone T09E11

C:Genetics:

A:Gene: CESP:T09E11.6

A:Map position: 1

A:Introns: 34/2; 54/2; 191/3; 275/3; 315/3; 359/3; 380/3; 403/3; 426/3; 530/1; 617/1; 66

Query Match

8.9%; Score 196; DB 2; Length 753;

Best Local Similarity 21.5%; Pred. No. 6.3e-07;

Matches 88; Conservative 72; Mismatches 195; Indels 54; Gaps 14;

```
QY 7 LVEXSL-----TSPFVRRRYTHVKEVR--YEVNCSGIYEOEP---LEIGKSLER 53
DB 366 LSESTLNCCEGPTKODLHHHTDIDADVREERGFADDRGSRRRPETAHVDCGRILAGD 425
QY 54 RRODIDLEDVVAMTSDCDIYOTLRGAKIVSKEEK-----SPLAYSLVYHKDAIMV 108
DB 426 KPYLOSITGTRKRVKIVENCNINNSCKAIRSRILPSNDNIIPLKHGIAFAFIYKDYEFI 485
QY 109 ERLIHAIYNQHNIIYCIHYDRKAPDPTFKVAMNNLAKCFSNIFIAKLEAVEYAHISRLQAD 168
DB 486 EKQYVSFHPQNAFCFVIDINASEEFKRRALACMPNVIYLADEDPVYSSGHNVLVH 545
QY 169 LNCISDLKSSIQMKYVINLCGQDFPLKSNPELVSELKINGANMLETVKPPNSKLERFT 228
DB 546 NKCKALKALIDIP-GNNYVALLLQNHDLIMKSYEMEQITEWLOGANDIFVT----- 593
QY 229 YHHELRRVPEYVKL-PI-----RTNISKAPPHNIQIFVGSAYFVLQAIFYKIENN 280
DB 594 --HEIGRDVAKKLKWDPMSTKLFTINETEMDKLLLTTPMKIYKGVWHSLSRASVEMMFOK 651
QY 281 SIQVDFE-AMSKDIYSPDEHFWATLIRP--GIRGEIS-RAQDVSDLOSRTLVKNNY 336
DB 652 LDPSTFEMQNLQGRGVDEQYFPIIQANAEFGMPGHHFTDCLQOGKTTERTTITIALM--- 708
QY 337 EGFFYP--SC-TGSHLRSCYIGAAELRWLIKDGHWFAKFDKSKVPDIL 382
DB 709 ----VPESKCDTMMTRHAVCIIGLEHFGAVASFTHLMFNKVSLSLDELI 753
```

Search completed: August 2, 2001, 14:45:51
Job time: 137 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: August 2, 2001, 14:44:50 ; Search time 34.61 Seconds
(without alignments)
726.927 Million cell updates/sec

```

Title: US-09-645-192-2_COPY_39_453
Perfect score: 2194
Sequence: 1 PQMDIYLVEYSLSTSPFVRN.....DWITPSEKLEMDRNLTTTS 415

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      412676 segs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```

post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database :

1. /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2. /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3. /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4. /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5. /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6. /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7. /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8. /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9. /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10. /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11. /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12. /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13. /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14. /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15. /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16. /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17. /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18. /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19. /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20. /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21. /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22. /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2194	100.0	453	22	AAU00037	Human CGRNT3, Homo
2	862	39.3	405	21	AAI18899	A core 2 beta-1,6
3	862	39.3	438	21	AAI18995	A core 2 beta-1,6
4	862	39.3	438	21	AAI94492	Human C2/4Gnt prot
5	862	39.3	438	22	AAI30518	Amino acid sequenc
6	862	39.3	465	21	AAI54344	Human pancreatic c
7	862	39.3	663	21	AAI18996	A partial core 2 b
8	854.5	38.9	428	19	AAW93942	Rat DHI protein.
9	849.5	38.7	428	19	AAW93943	Human core 2 Gnt E
10	848.5	38.7	428	21	AAI30208	Human core 2
11	840.5	38.3	428	15	AAI31366	Human heart core 2
						Sequence of human

12	838	38.2	427	21	AAB30927	Diabetic rat heart
13	711	32.4	400	16	AAR71932	I-branched enzyme
14	711	32.4	400	17	AA982474	Beta-1,6-N-acetylgl
15	711	32.4	400	19	AA956568	Full length amino
16	711	32.4	400	19	AAK47184	Human beta-1,6-N-a
17	383	17.5	126	17	AA932475	C2GnT catalytic do
18	355	16.2	126	19	AAW56627	Amino acid sequenc
19	355	16.2	126	19	AAW47185	Human beta-1,6-N-a
20	172	7.8	64	22	AAAB30519	Murine beta-1,6-N-a
21	166.5	7.6	426	21	AA9309038	Arabiidopsis thaliaa
22	166.5	7.6	434	21	AA9309037	Arabiidopsis thaliaa
23	158	7.2	334	21	AAAG90939	Arabiidopsis thaliaa
24	151	6.9	447	21	AAAG38142	Arabiidopsis thaliaa
25	148	6.7	486	21	AA938184	Arabiidopsis thaliaa
26	132	6.0	484	21	AA931854	Arabiidopsis thaliaa
27	132	6.0	630	21	AA931853	Arabiidopsis thaliaa
28	119	5.4	316	18	AAW14080	S.thermophilus exo
29	119	5.4	316	18	AAW22179	Arabiidopsis thaliaa
30	115	5.0	348	21	AAAG38143	Arabiidopsis thaliaa
31	109	5.0	289	21	AAAG37329	Arabiidopsis thaliaa
32	108	4.9	165	21	AAAG15071	Arabiidopsis thaliaa
33	107	4.9	328	21	AAAG38144	Arabiidopsis thaliaa
34	106	4.8	33	17	AA92475	C2GnT C-terminal r
35	104	4.7	2483	21	AAAB18172	Plasmodium falcipar
36	102.5	4.7	60	21	AAAG01848	Human secreted prote
37	102.5	4.7	726	12	AAAR10933	KGF receptor. Mus
38	102.5	4.7	726	12	AAAR14280	Murine KGF receptor
39	102.5	4.7	821	13	AAAR21080	Fly receptor proteor
40	102	4.6	707	21	AAAY95065	Candida albicans P
41	101.5	4.6	652	12	AAAR13269	Clone P7B183-encore
42	101	4.6	256	20	AAAY7314	Amino acid sequenc
43	101	4.6	285	21	AAAG47330	Arabiidopsis thaliaa
44	100	4.6	439	21	AAAY51247	C.roseus DAT protol
45	98.5	4.5	355	21	AAAG09800	Arabiidopsis thaliaa

ALIGNMENTS

RESULT

ID AAU00037 standard; Protein; 453 AA.

AC	AAU00037;
XX	
DT	14-MAY-2001 (first entry)

KW Human, 262513: Thymus-related disorder; cancer; tumour; adenoma;
 UDP-Glucose 6-phosphate dehydrogenase (NADP+ dependent) deficiency
 KW malignant melanoma; breast cancer; cervical cancer; hypacativity;
 KW hyperactivity; atrophy; Thymus enlargement; autoimmune; arthritis;
 KW leukaemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
 KW acquired immunodeficiency syndrome; sepsis; wound healing; infection.
 XX
 OS Homo sapiens.

FT	Key	Location/Qualifiers
FT	Domain	12..32
FT		/label= "Transmembrane domain"
FT	Protein	35..453
FT		/label= "Protein fragment retaining enzyme activity"
FT		/note= "used in a fusion protein for in vitro expression studies"
FT		72
FT	Modified-site	/note= "N-glycosylated"
FT		286
FT	Modified-site	/note= "N-glycosylated"
FT		317
FT	Modified-site	/note= "N-glycosylated"
FT		448
FT	Modified-site	/note= "N-glycosylated"

ET

XX MO200114535-AZ.
 PN
 XX
 PD 01-MAR-2001.
 XX
 XX 24-AUG-2000; 2000OWO-DK00469.
 PF
 XX 24-AUG-1999; 99US-0150488.
 PR
 XX (SCHM/) SCHMIENITEK T.
 PA (CLAU/) CLAUSEN H.
 XX
 PI Schwiientek T, Clausen H;
 DR WPI; 2001-226615/23.
 DR N-PSDB; AAS00045.
 XX
 PT New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
 PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
 PT thymus-related disorders, cancers, tumours, immunosuppression
 PS
 PS Claim 17; Fig 1; 97pp; English.

The sequence represents Human UDP-N-acetyl-glucosamine:Galactose-beta1,
 3-N-acetylglucosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
 (UDP-GlcNAc:Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
 C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
 preparation of compositions for treating a conditions mediated by C2GnT3,
 particularly a thymus-related disorder. C2GnT3, nucleic acids
 encoding it and antibodies against it may also be used for in vitro
 purposes related to scientific research, DNA synthesis and manufacture of
 vectors, in the prognostic and diagnostic evaluation of conditions
 associated with altered expression or activity of C2GnT3 or conditions
 requiring modulation of C2GnT3, as well as in monitoring conditions by
 detecting and localising the DNA and protein. Disorders such as tumours
 (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 breast or cervix), hypoactivity, hyperactivity, atrophy, enlargement of
 thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
 acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
 sepsis, wound healing, acute and chronic infection, cell-mediated or
 humoral immunity, or PH1/TH2 imbalance, may be treated using these protein
 or nucleic acid. The antibodies may be used to screen potential
 therapeutic compounds to determine their effects on a conditions such as
 thymus-related disorder or cancer, to determine the level of C2GnT3
 expression in cells genetically engineered to produce C2GnT3, or to
 detect and quantify polypeptides in a sample to determine their role in a
 particular cellular events or pathological states and to diagnose and
 treat such pathological states.

XX Sequence 453 AA;

Query Match 100.0%; Score 2194; DB 22; Length 453;
 Best Local Similarity 100.0%; Pred. No. 2,9e-190;
 Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 POKKITYLVESLSIPVYRNRTYVKEVYEVNCSGIYQEPLEIGKSLERRRIDL 60
 Db 39 PPKKITYLVESLSIPVYRNRTYVKEVYEVNCSGIYQEPLEIGKSLERRRIDL 98
 XX
 XX 61 EDDVYVMTSCDIDYQTLRGYAKLVSKESPEIAYSLVNHDAIWEHLIAIYQNH 120
 Db 99 EDDVYVMTSCDIDYQTLRGYAKLVSKESPEIAYSLVNHDAIWEHLIAIYQNH 158
 XX
 XX 121 IYCHYDRKAPDTRKVMNNIAKFSNITFASKIENVEYAHISLQADNLCLDLKST 180
 Db 159 IYCHYDRKAPDTRKVMNNIAKFSNITFASKIENVEYAHISLQADNLCLDLKST 218
 XX
 XX 181 QMKVYINLCGDPFLKSNFELVSELKLNANMLETVPKNSKLEERTYVHHELRVPEY 240
 Db 219 QMKVYINLCGDPFLKSNFELVSELKLNANMLETVPKNSKLEERTYVHHELRVPEY 278
 XX
 XX 241 VKLPRTNISKAPPHNIQIFVGSATYVLSQAFVKYIFNNSTYQDFANSKDTISPEDEH 300

Db 279 VKLPRTNISKAPPHNIQIFVGSATYVLSQAFVKYIFNNSTYQDFANSKDTISPEDEH 338
 XX
 XX 301 WATLIRVPGIGETISRSADVSDLOSKTRLVKNNYVGGFFYPSCTGSHLSVCIYAAEL 360
 Db 339 WATLIRVPGIGETISRSADVSDLOSKTRLVKNNYVGGFFYPSCTGSHLSVCIYAAEL 398
 XX
 XX 361 RWLIKDGHWFAKRPDSKYDPIILKICLAEKEBOORDMITLPSKLFMDRLTTTS 415
 Db 399 RWLIKDGHWFAKRPDSKYDPIILKICLAEKEBOORDMITLPSKLFMDRLTTTS 453
 XX

RESULT 2

ID AAB18999 standard; Peptide: 406 AA.

AC AAB18999;

DT 08-FEB-2001 (first entry)

DE A core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region.

XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW Cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis.

XX Homo sapiens.

XX CA2296936-A1.

XX 03-AUG-2000.

XX 03-FEB-2000; 2000CA-2296936.

XX 03-FEB-1999; 99US-0118674.

XX (GLYC-) GLYCDESIGN INC.

XX Korczak B, Lew A;

XX WPI; 2000-594746/57.

PT New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.

XX Claim 3; Page 53; 66pp; English.

The present sequence is derived from a human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.

XX Sequence 406 AA;

QY 188 LCGDPLKSNFELVSEIKLNGANMLETVKPNKSLERETVHHLRVRVEYKLPRT 247
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SO Sequence 465 AA:
 QY 13 STSPFYRN-RYTHVKDEYREYVNCSGIY--EQEPL--ETGKSLERRDITLEDDEYVA 67
 Db 74 sgqgcrnllynfiklpakrsincsgvtgqdeavlaqalinnlevkkr-efldthyls 132
 QY 68 MTSQCDIYQTLRGYAKLVSKSEKSPFIASLYVHKDAIWEELHAIYNQNHITVCHVD 127
 Db 133 ltrdcfhkaerikfidqfjskevefplaysmvlheklentferllavayapqnlvcvhvd 192
 QY 128 RKAPDTFKVAMNNLAKCFSNIFIAKSLAEVVAHISRLOADNCLSDLKSSIQMKRYVN 187
 Db 193 ekspeftkeavkaliscfpnvfiasklrvryaswsvrgdclmedllqssvwykfln 252
 QY 188 LCGDPLKSNFELVSEIKLNGANMLETVKPNKSLERETVHHLRVRVEYKLPRT 247
 Db 253 tcsqdfdkksnaemvqglkmlngnsesvppkhetrvkyfvevr---dlhl---t 306
 QY 248 NISKEAPPHNIQIFVGSAYFVLSQAFVYIFNNISIVODFFAMSKDYSPEDEHFWATLIRV 307
 Db 307 nkkdkpppylnltmtfgnayivasrdfvghvlpkpsqgllewvkdtyspdehlwatlqra 366
 QY 308 PGIPEEI-SRGAQDVSDLOSKTRLYKKNYEGF-----YPSCTGSLRVCYIGAAELR 361
 Db 340 rmpgsvpnbpkydisdmtslarlvkqghegdldkgapypapcsqllnqralvcygsqdln 399
 QY 362 WLKDGHWFAKFKDSKVDPIIKLAERL 390
 Db 400 wmlqnhllankfdpkyddnaqlcleyl 428
 RESULT 6
 AAB54344
 ID AAB54344 standard; Protein: 465 AA.
 XX
 AC AAB54344;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:796.
 XX
 KM Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KM detection; diagnosis; identification; cytostatic; neuroprotective;
 KM neoplastic; immunomodulatory; relaxant; contractile; gynaecological;
 KM antinflammatory; cardiac; gene therapy; chromosome mapping;
 KM linkage analysis; tissue identification; tissue typing; forensic;
 KM neural; immune system; muscular; reproductive; gastrointestinal;
 KM pulmonary; cardiovascular; renal; proliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO200055320-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05989.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-579444/54.
 DR N-PSDB; AAC99109.
 XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition
 XX
 PS Claim 11; Page 1237-1239; 1379pp; English.
 XX
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins. Called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neoplastic, immunomodulatory, relaxant, contractile,
 CC gynaecological, cardiac and antinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage

CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SO Sequence 465 AA:
 QY 13 STSPFYRN-RYTHVKDEYREYVNCSGIY--EQEPL--ETGKSLERRDITLEDDEYVA 67
 Db 74 sgqgcrnllynfiklpakrsincsgvtgqdeavlaqalinnlevkkr-efldthyls 132
 QY 68 MTSQCDIYQTLRGYAKLVSKSEKSPFIASLYVHKDAIWEELHAIYNQNHITVCHVD 127
 Db 133 ltrdcfhkaerikfidqfjskevefplaysmvlheklentferllavayapqnlvcvhvd 192
 QY 128 RKAPDTFKVAMNNLAKCFSNIFIAKSLAEVVAHISRLOADNCLSDLKSSIQMKRYVN 187
 Db 193 ekspeftkeavkaliscfpnvfiasklrvryaswsvrgdclmedllqssvwykfln 252
 QY 188 LCGDPLKSNFELVSEIKLNGANMLETVKPNKSLERETVHHLRVRVEYKLPRT 247
 Db 253 tcsqdfdkksnaemvqglkmlngnsesvppkhetrvkyfvevr---dlhl---t 306
 QY 248 NISKEAPPHNIQIFVGSAYFVLSQAFVYIFNNISIVODFFAMSKDYSPEDEHFWATLIRV 307
 Db 307 nkkdkpppylnltmtfgnayivasrdfvghvlpkpsqgllewvkdtyspdehlwatlqra 366
 QY 308 PGIPEEI-SRGAQDVSDLOSKTRLYKKNYEGF-----YPSCTGSLRVCYIGAAELR 361
 Db 340 rmpgsvpnbpkydisdmtslarlvkqghegdldkgapypapcsqllnqralvcygsqdln 399
 QY 362 WLKDGHWFAKFKDSKVDPIIKLAERL 390
 Db 427 wmlqnhllankfdpkyddnaqlcleyl 455
 RESULT 7
 AAB18996
 ID AAB18996 standard; Protein: 663 AA.
 XX
 AC AAB18996;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE A partial core 2 beta-1,6-N-acetylglucosaminyltransferase.
 XX
 KM Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KM cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KM rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KM septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KM platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KM clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KM diverticulitis; ulcerative colitis.
 XX
 OS Homo sapiens.
 XX
 PN CA2296936-A1.
 PN
 PD 03-AUG-2000.
 PD
 PF 03-FEB-2000; 2000CA-2296936.
 PF
 PR 03-FEB-1999; 99US-0118674.
 XX

PA (GLYC-) GLYCODESIGN INC.

XX Korczak B, Lew A;

XX WPI: 2000-594746/57.

XX New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.

PS Claim 3: Page 52; 66pp; English.

XX The present sequence represents a partial human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2 GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.

XX Sequence 663 AA:

Query Match 39.3%; Score 862; DB 21; Length 663;

Best Local Similarity 44.5%; Pred. No. 1.9e-69;

Matches 173; Conservative 70; Mismatches 128; Indels 18; Gaps 8;

QY 13 STSEFVN-RYTHVDEVRVNCSGY--EQEPL--EIKSLIRRRDIIJEDDDVYA 67

DB 131 sqsgyrcnllnyfklipakrslncosytrgdqavlaqlnmlevkkr-epftdhyis 189

QY 68 MTSCSDIYQRLRGYAKQVKEKSPFIAYSLVYHKDAIWERLIHAIVYNOHNYICHD 127

DB 190 ltrcencfkaerktqfpiskveefriaysmihetienfellerlrayapqilycwhd 249

QY 128 RKADDTFRVAMNNLAKCFNIFLASKLEAVEYAHISLQDLNCLSDLKSSIQWKYVIN 187

DB 250 ekspetkeavkaliscfpnvfiasklrvyvaswsvqadlnomedllqssvpmkyfln 309

QY 188 LCGDDPFLKSFELYSFLKLNQANMLQETVPPSKLERFTYHHELRVYEVYKLPRT 247

DB 310 tctgdtdfikaemvgaqiklmngnsmesepkpketrkynfevtr---delhl---t 363

QY 248 NISKEAPPNHIQIFVSGAYEVLQAFVYKIFNNSIVODEFPAWSKDTYSPDEHFWALIRV 307

DB 364 nkkddpppylntmtgnaivasrdvghvllknpksgqllewkdydspelhlwatqira 423

QY 308 PGIQGEI-SRKAQVSDLOKTRVYKKNYEGFR-----PSCGSLRSVCITYGAELR 361

DB 424 rmpggsvpnhpkysidmtstlarlvkwgngedldkxgapyapcghlqraivcygaadln 483

QY 362 WLIRGHWFAKFPDSKYVPIILIKLAERL 390

DB 484 wmlqnhhlankfdqpkvdnaqlgcleeyl 512

RESULT 8
 AAW93942
 ID AAW93942 standard; Protein; 428 AA.

XX AAW93942;

XX 28-JUN-1999 (first entry)

XX Rat Dhl protein.

XX Dhl; rat; screening; treatment; prevention; cardiomyopathy; inhibitor;
 KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
 KW UDP-GlcNAc:galbeta1-3galNAc-alphaR; transgenic animal; germ line;
 KW beta-1,6-N-acetylglucosaminyl-transferase.

XX Rattus sp.

XX CA2186987-A.

XX 02-APR-1998.

XX 02-OCT-1996; 96CA-2186987.

XX 02-OCT-1996; 96CA-2186987.

XX (MOUNT) MOUNT SINAI HOSPITAL CORP.

XX Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

XX WPI: 1998-399608/35.

XX N-PSDB; AAK24042.

PT Screening for substances that prevent or treat cardiomyopathy
 PT associated with diabetes and hyperglycaemia - comprises reacting
 PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
 PT in presence of test substance

PS Example 1; Fig 3A; 35pp; English.

XX This invention describes a method for screening for a substance that
 CC may be used to prevent or treat cardiomyopathy associated with diabetes
 CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
 CC (UDP-GlcNAc:galbeta1-3galNAc-alphaR beta-1,6-N-acetylglucosaminyl-
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in
 CC the presence of a test substance under conditions whereby the core 2
 CC GlcNAc-T produces a reaction product, determining the amount of reaction
 CC product, and comparing the amount of reaction product with the amount
 CC obtained in the absence of the test substance, where lower amounts of
 CC reaction product in the presence of the test substance indicate that the
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
 CC methods for preventing or treating cardiomyopathy associated with
 CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
 CC germ cells and somatic cells all contain a DNA construct introduced into
 CC the animal or an ancestor of the animal at an embryonic stage, where
 CC incorporation of the DNA construct into the germ line of the animal
 CC causes the animal to develop cardiomyopathy similar to that associated
 CC with diabetes mellitus and hyperglycaemia. This sequence represents the
 CC rat Dhl protein which is used to describe the method of the invention.

XX Sequence 428 AA:

Query Match 38.9%; Score 854.5; DB 19; Length 428;

Best Local Similarity 42.1%; Pred. No. 4.8e-69;

Matches 177; Conservative 71; Mismatches 133; Indels 39; Gaps 10;

QY 1 POKDITLY-EYSLSTSPFVRNR-----YTHVK--DEVREYVNCSGIYEOEPLKIK- 48

DB 14 pckyytmvavlslltsvtrlnqkpefvsvshlsgddnsvnrcvklvgdpeelgkv 73

QY 49 SLEI-----RRRDIILEDDDVVAMTSCDIIYQTLRGYAKQVKEKSPFIAYSL 99

DB 74 kleiltvgfkxprtr-----cpdylnmtrcaasfirtkyimpeklkeeygfplayst 127

QY 100 VYHKDAIWERLIHAIVYNOHNYICHIYDKRADDFKVAAMNNLAKCFNIFLASKLEAVEY 159

DB 128 vvhkxldmldrlralympqnfycihvdrkaeesflaavgiaascfdnvfasqlesvvy 187

QY 160 AHISRLQADLNCISDLKSSIQWKYVINLCGDDPFLKSFELYSFLKLNQANMLQETVPP 219

Db 188 aswsvkAdIncmkdlYrmanwkylinlcmgdfpikntleivrklsftgensletekm 247
 QY 220 PMSKLEPFTYHHELRPRYEVYKLPRTNISKAPPHNIQIFVGSAYFLVSOAFVXYIFN 279
 Db 248 pnmkeewk-----krytvdgkl-tngyvkagpriktpifsgsfvvtreyvgvyle 301
 QY 280 NSIVQDFEAMSKDTPSPDEHFNATLIRVPGIGEISRSQA-DVSDLOSKTRLYKMNYYEG 338
 Db 302 nkniqkfmewagdtyspdefiwtatqirpevgslpsshkydlsdmavartfwkwyfeg 361
 QY 339 FF-----YPSCTGSHLSVCIYGAELRWLIDGHWANFPDSKVDPILIKCAEKLIEQ 393
 Db 362 dvsngapyppcgsvhvtrsvclfgvgdlsmlrkhnfrankfmdvdpfaigclseelnrhk 421
 RESULT 9
 AAM93943
 ID AAM93943 standard; Protein: 428 AA.
 XX
 AC AAM93943;
 DT 28-JUN-1999 (first entry)
 XX
 DE Human core 2 Gnt protein.
 XX
 KW Screening; treatment; prevention; cardiomyopathy; inhibitor;
 RW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
 KM UDP-GlcNAc:Galbeta1-3GalNAc-alpha; transgenic animal; germ line;
 XX beta-1,6-N-acetylglucosaminyl-transferase; human; core 2 Gnt.
 OS Homo sapiens.
 XX
 PN CA2186987-A.
 XX
 PD 02-APR-1998.
 XX
 PF 02-OCT-1996; 96CA-2186987.
 XX
 PR 02-OCT-1996; 96CA-2186987.
 XX
 PA (MOUN) MOUN SINAI HOSPITAL CORP.
 XX
 PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;
 DR MPI: 1998-399608/35.
 DR N-PSDB; AAX24043.
 XX
 PT Screening for substances that prevent or treat cardiomyopathy
 PT associated with diabetes and hyperglycaemia - comprises reacting
 PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
 PT in presence of test substance
 XX
 PS Disclosure; Fig 9; 35pp; English.
 CC This invention describes a method for screening for a substance that
 CC may be used to prevent or treat cardiomyopathy associated with diabetes
 CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
 CC (UDP-GlcNAc:Galbeta1-3GalNAc-alpha; beta-1,6-N-acetylglucosaminyl-
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in
 CC the presence of a test substance under conditions whereby the core 2
 CC GlcNAc-T produces a reaction product, determining the amount of reaction
 CC product, and comparing the amount of reaction product with the amount
 CC obtained in the absence of the test substance, where lower amounts of
 CC reaction product in the presence of the test substance indicate that the
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
 CC methods for preventing or treating cardiomyopathy associated with
 CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
 CC germ cells and somatic cells all contain a DNA construct introduced into
 CC the animal or an ancestor of the animal at an embryonic stage, where
 CC incorporation of the DNA construct into the germ line of the animal
 CC causes the animal to develop cardiomyopathy similar to that associated
 CC with diabetes mellitus and hyperglycaemia. This sequence represents the
 CC human core 2 Gnt protein (beta-1,6-N-acetylglucosaminyltransferase)

CC which is used to describe the method of the invention.

XX SQ Sequence 428 AA;

Query Match 38.7%; Score 849.5; DB 19; Length 428;
 Best Local Similarity 43.0%; Pred. No. 1,4e-68;
 Matches 181; Conservative 70; Mismatches 129; Indels 41; Gaps 11.

QY 1 POKDIYV-EYSLSTSPFVFNRYTHVKNDE---VRY-----EVNCSGIYDEPLEI 46
 Db 14 pckyyfmvvlvslstfsvlrl---lbqkpefsvvrhlslagengssdinctkvlggdvnei 70
 QY 47 GK-----SEIFRRDIIIDEDDDVYAMTSDCDIYOTLGRYAKQIYVSKESKSEPIAYSLV 100
 Db 71 qkvkllllvklfkkrp--twcpddylnmtdscsstikrktylvepiskseaeafplaysiv 128
 QY 101 VHKDAIINVERLIRAIYHQNHYCIHYDRKAPDTFFKYAMNNLAKCSNIFIAKLEAVEYA 160
 Db 129 vhhklemldrlrlaiympgnfyvvhvdtksedsylaavmgiascfenvfvasrlseavya 188
 QY 161 HISRLQADNLCLSDLKSSIQMKYVINLCGDPPLKSNFELVSELKILNGAMMLETVKRP 220
 Db 189 swarvgadlncmkdlYamsanwkylinlcmgdfpikntleivrkiklmgennletermp 248
 QY 221 NSKLERFTYHHELRPRYEVYK-LPRTNISKAPPHNIQIFVGSAYFLVSOAFVXYIF 278
 Db 249 shkeerw-----kryevvngkl-tngtvmmlpplletplsgsagylvtrsvreygyl 300
 QY 279 NSIVQDFEAMSKDTPSPDEHFNATLIRVPGIGEISRSQA-DVSDLOSKTRLYKMNYYE 337
 Db 301 qneikqlmewagdtyspdeylwatlqirpevgslpsshkydlsdmavartfwkwyfeg 360
 QY 338 GFF-----YPSCTGSHLSVCIYGAELRWLIDGHWANFPDSKVDPILIKCAEKLIEQ 392
 Db 361 gdwskgapyppcdgsvhvtrsvclfgvgdlnmrlkhnfrankfmdvdpfaigclseelnrhk 420
 QY 393 Q 393
 Db 421 k 421
 RESULT 10
 AAB30298
 ID AAB30298 standard; Protein: 428 AA.
 XX
 AC AAB30298;
 DT 12-FEB-2001 (first entry)
 XX
 DE Human heart core 2 GlcNAc-T.
 XX
 KW Human; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
 KW UDP-GlcNAc:Galbeta1-3GalNAc-alpha; beta-1,6-N-acetylglucosaminyltransferase.
 XX
 OS Homo sapiens.
 XX
 PN US6131578-A.
 XX
 PD 17-OCT-2000.
 XX
 PF 02-OCT-1997; 97US-0943058.
 XX
 PR 02-OCT-1996; 96US-0046876.
 XX
 PA (KING/) KING G L.
 PA (NISHU/) NISHIO Y.
 PA (KOYA/) KOYA D.
 PA (DENNI/) DENNIS J W.
 PA (WARRE/) WARREN C E.
 XX
 PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;

XX WPI: 2000-678642/66.
 DR N-PSDB; AAC65469.
 XX
 PT Preventing or treating cardiomyopathy associated with diabetes mellitus
 PT and hyperglycemia, comprises administering a substance that inhibits
 PT core 2 N-acetylglucosamine-T activity -
 PS
 PS Claim 3; Fig 9; 21pp; English.
 CC The present invention describes the human UDP-glucNAc:galbeta-1-3galNAc
 CC alpha/beta 6-N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
 CC and coding sequences. The enzyme is associated with cardiomyopathy in
 CC diabetes and hyperglycemia sufferers. The invention also provides
 CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
 CC in treatment.
 CC
 XX
 XX Sequence 428 AA:
 SO
 Query Match 38.7%; Score 849.5; DB 21; Length 428;
 Best Local Similarity 43.0%; Pred. No. 1.4e-68;
 Matches 181; Conservative 70; Mismatches 129; Indels 41; Gaps 11;
 QY 1 POKDIYLV-EYSISTSPFVNRNRYTHWKDE---VRY-----EVNCSGIYGEPELEI 46
 DB 14 ptkytmvnlvsltsfsvlr---lhqkpefsvrnhlelagempssdlnctkylgqdvnei 70
 QY 47 GK-----SLEIRRRDIDLEDDVYAMTSCDIYOTLRGYAKLVSKEEKSPFIAYSLV 100
 DB 71 qkvlelltvkfkfkrp--twppddylnmtsdcsfklrpxylveplskaeefrlaysiv 128
 QY 101 VHKDAIWEELHAIYQNHNYICIHDRKAPDFEKYAMNNLACFSNFIASFLAEVEXA 160
 DB 129 vnhklemldrlsaiyipgnfyvhdtkseedsyaaavgiascfsnvfvaarlesvyya 188
 QY 161 HISRLQADNLCLSDLKSSIQWKYVYNLCGQDFPLKSNFELVSELKLNANGAMLETVPK 220
 DB 189 swsrvgadlnckmdlyamsanwkyllnlgcmdfpikntleivrkiklimgennletemp 248
 QY 221 NSKLERETYNHELRRVPEYEV--KLPIRTNISKAPPHNIQIFVGSAYFVLSQAFAKYIF 278
 DB 249 shkeerw-----kkrtyevvngkl-tngtvtmblpplctplfsagsayfvsvreygyvl 300
 QY 279 NNSIVODFFAMSKDYSPEDEHFNATLIRVPGIPEISRSAQ-DVSDLSQKTRLVKNNYIE 337
 DB 301 qnekiqlmewagdytspdeylwatigripevpslpashkydlisdmgavarfkwqyfe 360
 QY 338 GFF-----YPSCTGSHLRSCVCIYGAELRWLIDGHWFAFKPDSKVPILIKLAEKLEE 392
 DB 361 gdvskgapyppcdgwhvsvcltfgagdlmwmlrkhlfnakfdvadvdlfaicqidenhln 420
 QY 393 Q 393
 DB 421 k 421
 RESULT 11
 AARS1386 standard; Protein; 428 AA.
 XX AARS1386;
 AC
 XX
 XX 07-OCT-1994 (first entry)
 DE Sequence of human core 2 beta 1-6
 DE N-acetylglucosaminyltransferase (C2GNT or core 1-6 AGT).
 DE
 DE C2GNT; 1-6 AGT; core 2 beta 1-6 N-acetylglucosaminyltransferase;
 KM O-glycan.
 XX
 XX Homo sapiens.
 OS
 XX

EH Key Location/Qualifiers
 FT Domain 10..32
 FT 10..32
 FT Modified-site 52
 FT 52
 FT /label= potential N-glycosylation site
 FT /note= "see also AAs 58 and 95"
 XX
 XX Eps90747-A.
 XX
 XX 06-APR-1994.
 XX
 XX 29-SEP-1993; 93EP-0250268.
 XX
 XX 01-OCT-1992; 92US-0955041.
 XX
 XX (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 XX Biehuizen MFA, Fukuda M;
 XX
 XX WPI: 1994-111195/14.
 DR N-PSDB; AA061559.
 XX
 XX New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are
 FT used for the study of the effect of variant O-glycan(s) on
 PT cell-cell interactions, partic. in cancers
 PS
 PS Disclosure; Page 20-22; 34pp; English.
 CC C2GNT, or an active fragment thereof, catalyses the formation of
 CC critical branches in O-glycans. cDNAs encoding various
 CC glycosyltransferases can be isolated by transient expression of cDNA
 CC in recipient cells, e.g. COS-1. COS-1 cells were transfected
 CC with a cDNA library, pCDR alpha-2R1, constructed from poly(A)+ RNA
 CC of activated T lymphocytes which express the C2GNT. Transfected
 CC cells were selected using Mab T305, which identifies a
 CC hexasaccharide on leukostatin. Leukostatin CD43 is an acceptor
 CC molecule for C2GNT activity. A plasmid, pCDR alpha-leu,
 CC which directed expression of the T305 antigen was identified. The
 CC cDNA insert was isolated and sequenced. The cDNA encoded the
 CC acceptor molecule leukostatin CD43.
 CC
 XX
 XX Sequence 428 AA:
 SO
 Query Match 38.3%; Score 840.5; DB 15; Length 428;
 Best Local Similarity 42.8%; Pred. No. 8.8e-68;
 Matches 180; Conservative 70; Mismatches 130; Indels 41; Gaps 11;
 QY 1 POKDIYLV-EYSISTSPFVNRNRYTHWKDE---VRY-----EVNCSGIYGEPELEI 46
 DB 14 ptkytmvnlvsltsfsvlr---lhqkpefsvrnhlelagempssdlnctkylgqdvnei 70
 QY 47 GK-----SLEIRRRDIDLEDDVYAMTSCDIYOTLRGYAKLVSKEEKSPFIAYSLV 100
 DB 71 qkvlelltvkfkfkrp--twppddylnmtsdcsfklrpxylveplskaeefrlaysiv 128
 QY 101 VHKDAIWEELHAIYQNHNYICIHDRKAPDFEKYAMNNLACFSNFIASFLAEVEXA 160
 DB 129 vnhklemldrlsaiyipgnfyvhdtkseedsyaaavgiascfsnvfvaarlesvyya 188
 QY 161 HISRLQADNLCLSDLKSSIQWKYVYNLCGQDFPLKSNFELVSELKLNANGAMLETVPK 220
 DB 189 swsrvgadlnckmdlyamsanwkyllnlgcmdfpikntleivrkiklimgennletemp 248
 QY 221 NSKLERETYNHELRRVPEYEV--KLPIRTNISKAPPHNIQIFVGSAYFVLSQAFAKYIF 278
 DB 249 shkeerw-----kkrtyevvngkl-tngtvtmblpplctplfsagsayfvsvreygyvl 300
 QY 279 NNSIVODFFAMSKDYSPEDEHFNATLIRVPGIPEISRSAQ-DVSDLSQKTRLVKNNYIE 337
 DB 301 qnekiqlmewagdytspdeylwatigripevpslpashkydlisdmgavarfkwqyfe 360
 QY 338 GFF-----YPSCTGSHLRSCVCIYGAELRWLIDGHWFAFKPDSKVPILIKLAEKLEE 392

Db 361 gatskagapppcdgvyhvsvcifegdlmmlrknhlfavdvdlfaicldentrh 420
 QY 393 Q 393
 Db 421 k 421

RESULT 12

AAB30297
 ID AAB30297 standard; Protein; 427 AA.

AC AAB30297;

DT 12-FEB-2001 (first entry)

DE Diabetic rat heart core 2 GlcNAc-T.

XX Rat, diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
 KW UDP-GlcNAc:Galbeta1-3GalNAcAlphar beta1-6
 KM N-acetylglucosaminyltransferase.

XX Rattus sp.

PN US6131578-A.

PD 17-OCT-2000.

PF 02-OCT-1997; 97US-0943058.

PR 02-OCT-1996; 96US-0046876.

PA (KING/) KING G L.

PA (NISH/) NISHIO Y.

PA (KORA/) KORA D.

PA (DENNIS/) DENNIS J W.

PA (WARR/) WARREN C E.

PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;

DR WPI: 2000-678642/66.

DR N-PSDB: AAC65468.

XX Preventing or treating cardiomyopathy associated with diabetes mellitus

PT and hyperglycemia, comprises administering a substance that inhibits

PT core 2 N-acetylglucosamine-T activity.

XX Example 1; Fig 3B; 21pp; English.

XX The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc

CC alphas beta1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein

CC and coding sequences. The enzyme is associated with cardiomyopathy in

CC diabetes and hyperglycaemia sufferers. The invention also provides

CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used

CC in treatment.

XX Sequence 427 AA;

SO

Query Match 38.2%; Score 838; DB 21; Length 427;

Best Local Similarity 40.8%; Pred. NO. 1.5e-67;

Matches 172; Conservative 76; Mismatches 130; Indels 44; Gaps 11;

QY 1 POKDIIV-EYSLSTSPFRRN-----YTHVK---DEVREVNCSGIYQOELEIGKS 49

Db 14 pckymfmlvslitfevrltqkpefsvshleibgdpsnnctkylgdpdeeqkv 73

QY 50 LEIRRDIDLED-----DDVYAMSDCDIYOTLFGYKQKLVSKREKSPFIATSLV 101

Db 74 ---Kxlltvgckrrpttphetumlrtdcasiflrltkymepfkeeygfplaysiv 129

QY 102 HKDAIVVERLHAIVNOHNYCIHYDRKAPDTFKVAMNNLAKCSNIPFIASKLAVEYAH 161

Db 130 hkhidmldtllialympqatfycihvdtrkaeesflaavvgiascfduvfasqlesvvyas 189
 QY 162 ISRLQADNLCLSDLLKSSIQMKRYVINLCGDPPLKSNFEIVSEKLTGANMETVX-PP 220
 Db 190 wsvxkadlncmkdlyrmanvkylinlcmgdfplktlnleivrkllsfvgsnletekmp 249
 QY 221 NSKLERFTHHLERRVPEYV---KLPIRTNISKAPPHNIQIFVGSAYEVLQAFVKY 277
 Db 250 nke-----krwkyzyvvdgkl-tnlyvrxagpplktpilsgasayfvttreygyv 298
 QY 278 FNNISIVQDFPANSKDTYSPDEHFWATLIRVPEIGELISRSQ-DVSDLSQKTRLVKNY 336
 Db 299 lenkniqkimewagdtyspdehfwatlgirdey9slpsshkyldsdmnavarfxwqif 358
 QY 337 EGFPE-----YPSCTGSHLRVSVCIYGAELRWLIKDGHEWFAKFPDSKVPILIKLAETK 391
 Db 359 egdvsngagppcsgvyhvsvcifvgdlswmrlrhlfankfmdvdpfalqleehlr 418
 QY 392 EQ 393
 Db 419 hk 420

RESULT 13

AAR71932
 ID AAR71932 standard; Protein; 400 AA.

AC AAR71932;

DT 23-SEP-1995 (first entry)

DE I-branching enzyme.

XX Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; Ignt;

KW inflammation; antitumorigenic; tumor; hypersensitivity; anemia;

KW transgenic animal; I antigen.

XX Homo sapiens.

PN WO9507020-A.

PD 16-MAR-1995.

PF 09-SEP-1993; 93WO-US08476.

PR 09-SEP-1993; 93WO-US08476.

PA (LJOL-) LA JOLLA CANCER RES FOUND.

PI Bierhuizen MFA, Fukuda M;

DR WPI: 1995-123182/16.

DR N-PSDB: AA089201.

XX Nucleic acid encoding I-branching enzyme - used to develop prods.

PT for treating e.g. inflammatory responses, tumours or

PT hypersensitivity reactions

PS Disclosure: Fig.6; 56pp; English.

XX Poly-A RNA isolated from human PA-1 teratocarcinoma cells, was

CC reverse-transcribed and inserted into expression vector pCDNA1.

CC Plasmid DNA was used to transfect CHO-Py-leu cells. Transfected

CC cells were screened using human anti-I antigen antibodies and goat

CC anti-human IgM. Plasmid DNA was analysed to obtain DNA (given in

CC AA089201) encoding I-branching enzyme (AAR71932).

SO Sequence 400 AA;

Query Match 32.4%; Score 711; DB 16; Length 400;

Best Local Similarity 43.7%; Pred. NO. 4.2e-56;

Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

XX W0200114535-A2.
 PN
 XX
 XX 01-MAR-2001.
 PD
 XX
 XX 24-AUG-2000; 2000WO-DK00469.
 PF
 XX
 XX 24-AUG-1999; 99US-0150488.
 PR
 XX
 XX (SCHW/) SCHWIENTEK T.
 PA (CLAU/) CLAUSEN H.
 XX
 XX Schwianteck T, Clausen H;
 PI
 XX WPI: 2001-226615/23.
 DR N-PSDB; AAS00045.
 XX
 XX New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
 PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
 PT thymus-related disorders, cancers, tumours, immunosuppression
 XX
 XX Claim 17, Fig 1, 97pp: English.
 PS
 XX The sequence represents Human UDP-N-acetyl-glucosamine:Galactose-beta1,
 CC 3-N-acetylglucosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
 CC (UDP-GlcNAc:Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
 CC C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
 CC preparation of compositions for treating a conditions mediated by C2GnT3,
 CC particularly a thymus-related disorder. C2GnT3, nucleic acids
 CC encoding it and antibodies against it may also be used for in vitro
 CC purposes related to scientific research, DNA synthesis and manufacture of
 CC vectors, in the prognostic and diagnostic evaluation of conditions
 CC associated with altered expression or activity of C2GnT3 or conditions
 CC requiring modulation of C2GnT3, as well as in monitoring conditions by
 CC detecting and localising the DNA and protein. Disorders such as tumours
 CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 CC breast or cervix), hypoactivity, hyperactivity, atrophy, enlargement of
 CC thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
 CC acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
 CC sepsis, wound healing, acute and chronic infection, cell-mediated or
 CC humoral immunity, or Th1/Th2 imbalance, may be treated using these protein
 CC or nucleic acid. The antibodies may be used to screen potential
 CC therapeutic compounds to determine their effects on a conditions such as
 CC thymus-related disorder or cancer, to determine the level of C2GnT3
 CC expression in cells genetically engineered to produce C2GnT3, or to
 CC detect and quantify polypeptides in a sample to determine their role in a
 CC particular cellular events or pathological states and to diagnose and
 CC treat such pathological states.
 CC
 XX
 XX Sequence 453 AA;
 SQ
 Query Match 100.0%; Score 2389; DB 22; Length 453;
 Best Local Similarity 100.0%; Pred. No. 1.6e-207;
 Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 241 selklingannmlekvkpnsklerftlyhbelrrrpyevkrlptntnlskeepn:qflv 300
 |||
 QY 301 GSAYFVLQAFVKYIFENNSIVQDFPFAWSKDTYSPDEHFWALLIVPGIPGITSADVS 360
 |||
 DB 301 .gsayfvlsgaivkylfnnsvqdfifawskdypdehfwallivpgipgitsadvs 360
 |||
 QY 361 DQSKTRLVKNWYVYEGFYPSCGSHLSVCITYGAELRWLIKGDHWFANFDSKVPIL 420
 |||
 DB 361 dqsctrlvknwnyvgfypscgshlsvcitygaelrwlkdgwhfanfdsckvpil 420
 |||
 QY 421 IKCLAEKLEEQORDWITLPSEKLEMDRNLTTS 453
 |||
 DB 421 ikclaekeeqrdwiltipseklmdrnlts 453
 |||
 RESULT 2
 AAB18995
 ID AAB18995 standard; Protein: 438 AA.
 XX
 AC AAB18995;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 XX A core 2 beta-1,6-N-acetylglucosaminyltransferase.
 DE
 XX
 KW Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; AIDS; cancer;
 KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis.
 KW
 OS Homo sapiens.
 XX
 CA2296936-A1.
 PN
 PN 03-AUG-2000.
 PD
 PD 03-FEB-2000; 2000CA-2296936.
 PE
 PE 03-FEB-2000; 2000CA-2296936.
 XX
 PR 03-FEB-1999; 99US-0118674.
 XX
 PA (GLYC-) GLYCDESIGN INC.
 XX
 PI Korcek B, Lew A;
 PI
 XX
 XX WPI: 2000-594746/57.
 DR N-PSDB; AAA96569.
 DR
 XX
 PT New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 XX
 XX Claim 3; Page 50-51; 66pp; English.
 PS
 XX The present sequence represents a human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular

XX	06-MAR-2001	(first entry)
DE	Amino acid sequence of beta-1-6-N-acetylglucosaminyltransferase.	
KM	Human: beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M; inflammation;	
KW	membrane protein; branched sialyl Lex; L-selectin; immune reaction;	
RW	Inflammation; tissue rejection; tumour metastasis.	
OS	Homo sapiens.	
FN		
FT	Key Location/Qualifiers	
FT	Modified-site 289	
XX	/note= "potential N-glycosylation site"	
XX	US6136580-A.	
PD	24-OCT-2000.	
PE	19-JAN-1999; 99US-0233506.	
PR	19-JAN-1999; 99US-0233506.	
PA	(BURN-) BURNEM INST.	
P1	Fukuda M, Yeh J;	
DR	WPI; 2001-040238/05.	
XX	N-PSDB; AAC62134.	
PT	New C2GNT-M polypeptides having core 2, core 4 and I branching	
PT	beta-1-6-N-acetylglucosaminyltransferase activities for preparing	
PT	reagents useful for diagnosing, preventing or treating inflammation or	
XT	tumour metastasis -	
PS	Claim 1; Fig 4; 25pp; English.	
CC	The present sequence represents a human	
CC	beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4	
CC	and I branching activities. It is designated C2GNT-M. C2GNT-M is a	
CC	membrane protein that is predominantly expressed in colon, small	
CC	intestine, trachea, stomach and thyroid, as well as in certain cancer	
CC	cell lines. C2GNT-M polypeptides may be used to prepare molecules having	
CC	highly branched sialyl Lex and L-selectins, which may be subsequently	
CC	used to modulate immune reactions, e.g. inflammation and tissue	
CC	rejection, and to prevent or inhibit tumour metastasis.	
SQ	Sequence 438 AA:	
	Query Match 36.3%; Score 868; DB 22; Length 438;	
	Best Local Similarity 43.0%; Pred. No. 4,4e-70f;	
	Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps	
OY	23 LMLSLKLLNV-----RLEFPQKDILVEYSLSGPSFVRN RTHVHKDEVREYVNCSG 75	
DB	13 lvalgsgymlatlvaalklsfrfkcdshlglesresgsqycrnlllyflilipakrsincsg 72	
OY	76 IV--EDELPL--EIGKSLEIRBRDIDCEDDDVYAMSDCDIYOTLNGYAKOKLVSKDEKSF 131	
DB	73 vctgdgeavlgdalmlnlekkkr--efldthyslrlcdehfhaekrkffiqfpkskevef 131	
OY	132 PLAYSLVHVKDAIMVERLIHAIFYNQHNICYIHDKRAPTDFKYAAMNNLAKCSNFITASK 191	
DB	132 playsmvlhekienferillrtavyapqnlycvhvdxespetfxkaavkalisdpnvitask 191	
OY	192 LEAVEFAHSRLQADNLCLSDLKSSIQWKYYVNLCGODPFLKSNFELYSELKKLGANN 251	
DB	192 lrvvyvasvstrgadlnmedilqssvpmkyflntcgtdfpiknsaemayaklmIngns 251	
OY	252 LETVKKPPNSGLEFFTYHHELRRAPRYEVKKRPITNTINSKEAPRNIIQFVSATFYLSQA 311	
DB	252 meevvpkhketrwkyhlfevr--dlhnl--tnkkkappynlltmftgnayivaardf 305	

OY WKLENNISVODFFRAMSKNDTSPENHMAITLIRVPJIGET-SSAQDVSPLOSTKTLVK 370
312 :|||
Dd vqghlkmprksqqlliewdydtyspenhlwtqlqrmpgsyvpnlpxkdlsdmstiarlvk 365
306 :|||||
OY 371 WNYEGGF-----YPCTGSHLRVCVIYGAAELRWLIKDGHWPAKNRDSKYDPIILIKCLA 425
370 :|||
Dd wqbhgagdidkgapypcsglhqraicvygaagdlnmmlqnhlhlanfkfdpkvddnaiqcle 425
366 :|||
OY 426 EKL 428
426 ey1 428
Dd

RESULT 5
AAB18996
ID AAB18996 standard; Protein: 663 AA.

AAB18996;
08-FEB-2001 (first entry)

A partial core 2 beta-1,6-N-acetylglucosaminyltransferase.
Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLCNAC-T;
cancer: cardiovascular disorder; inflammatory disorder; asthma;
rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
septic shock; adult respiratory distress syndrome; ARDS; cancer;
platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
diverticulitis; ulcerative colitis.
Homo sapiens.
OS Homo sapiens.
XX
PN CA2296936-A1.
XX
PD 03-AUG-2000.
XX
PF 03-FEB-2000; 2000CA-2296936.
XX
PR 03-FEB-1999; 99US-0118674.
XX
PA (GLYC-) GLYCDESIGN INC.
XX
PI Korczak B, Lew A;
XX
DR MPI; 2000-594746/57.
XX
PT New nucleic acid molecules of core 2
XX
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
compositions for treatment of disorders mediated by the enzyme
including cancer, cardiovascular and inflammatory disorders.
PS Claim 3; Page 52; 66pp; English.
XX

The present sequence represents a partial human core 2
beta-1,6-N-acetylglucosaminyltransferase (core2b GLCNAC-T) polypeptide.
The polypeptide can be used to treat diseases and disorders, such as
cancer, cardiovascular disorders and inflammatory disorders including
asthma, rheumatoid arthritis, inflammatory bowel disease,
arteriolosclerosis, septic shock, adult respiratory distress syndrome
(ARDS), and cancer. Various platelet-mediated pathologies such as
atherosclerosis and clotting can also be treated. The polypeptides of
the invention are predominantly expressed in gastrointestinal tissue
(stomach, colon, intestine, testis) and are elevated in cancer.
Gastrointestinal disorders that may be prevented or treated include
ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
and ulcerative colitis. The antibodies may be used in
immunohistochemical analysis, to detect the novel polypeptide and to
localize it to particular cells and tissues and to specific subcellular
locations and to quantitate the level of expression.

Sequence 663 AA:

Query Match 36.3%; Score 868; DB 21; Length 663;
 Best Local Similarity 43.0%; Pred. No. 8.1e-70;
 Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;

```

QY 23 LMLSLKLKLVN-----RRLEPKDILYVYSLSTSPFVRN-RYTHVKNDEVRVENCSCG 75
DB 97 lwaigcmllatvalklstfrkcdsdhglstresgsgcynlllynfiklpakstincsg 156
QY 76 IY--EOEPL--EIGKSEIRRRDIIDDDPVAMTSDCDIYQTLRGYAQLVSKDEKSF 131
DB 157 vtrgdeevlqalnnlevkkkr-epfdthylstldcehfkaerfifqfplskseevf 215
QY 132 FIATSLVYHKAIMVERLIHAIYQNHNYCYCHYDRKAPDTFKVAMNNLANCFSTFIASK 191
DB 216 playsmvthekienferlliravapqnlcyvhvdekspefkkeavkaliscfpvltask 275
QY 192 LEAVEYAHISRLOADLNCISDLKSSIQWKYVNLGQDPFLKSNFELVSELKXLNGAMM 251
DB 276 lrvvyaswstrvqadlnmedlqsgvpmkyflntcgdfpikanaemvgalkmlngns 335
QY 252 LETVAPPSKLERFTYHHELRVPEYVKLPRTINISKEAPPNIQIFVGSATVLSQAF 311
DB 336 mesevppkhetkrkyhfevvr---dlhl---lnkkkdpplnltfignaylvaaardf 389
QY 312 VKYIFNNSIVODEFAMSKDTYSPDEHFWATLIRVGPGEI-SRSADVDSDLOSKTRLVK 370
DB 390 vghvlnkpksgqllewvkdtyspdehfwatlgqrarmpgsvpnphkylsdmslarlvk 449
QY 371 WNYVGEFF----YPSCTGSHLSVCIYGAELRWLKGHWFRANKFDSKVDPLIKCLA 425
DB 450 wqghgedldkqapypcsglbgtralcvygagdlmmlqnhllanktkpkrvdnaqlcle 509
QY 426 EKL 428
DB 510 eyl 512

RESULT 6
AAM93942
ID AAM93942 standard; Protein; 428 AA.
XX
AC AAM93942;
XX
DT 28-JUN-1999 (first entry)
XX
DE Rat DH1 protein.
XX
KM DH1; rat; screening; treatment; prevention; cardiomyopathy; inhibitor;
KM diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
KM UDP-GlcNAc:Galbeta1-3GalNAc-alphaR; transgenic animal; germ line;
KM beta-1,6-N-acetylglucosaminyl-transferase.
XX
OS Rattus sp.
XX
CA2186987-A.
XX
PN CA2186987-A.
XX
PD 02-APR-1998.
XX
PF 02-OCT-1996; 96CA-2186987.
XX
PR 02-OCT-1996; 96CA-2186987.
XX
PA (MOUN ) MOUNT SINAI HOSPITAL CORP.
XX
PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;
XX
DR WPI; 1998-399608/35.
XX
DR N-PSDB; AAX24042.
XX
PT Screening for substances that prevent or treat cardiomyopathy
PT associated with diabetes and hyperglycaemia - comprises reacting
PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
PT
```

PT in presence of test substance

PS Example 1; Fig 3A; 35pp; English.

CC This invention describes a method for screening for a substance that
 CC may be used to prevent or treat cardiomyopathy associated with diabetes
 CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
 CC (UDP-GlcNAc:Galbeta1-3GalNAc-alphaR beta-1,6-N-acetylglucosaminyl-
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in
 CC the presence of a test substance under conditions whereby the core 2
 CC GlcNAc-T produces a reaction product, determining the amount of reaction
 CC product, and comparing the amount of reaction product with the amount
 CC of reaction product in the absence of the test substance, where lower amounts of
 CC reaction product in the presence of the test substance indicate that the
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
 CC methods for preventing or treating cardiomyopathy associated with
 CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
 CC germ cells and somatic cells all contain a DNA construct introduced into
 CC the animal or an ancestor of the animal at an embryonic stage, where
 CC incorporation of the DNA construct into the germ line of the animal
 CC causes the animal to develop cardiomyopathy similar to that associated
 CC with diabetes mellitus and hyperglycaemia. This sequence represents the
 CC rat DH1 protein which is used to describe the method of the invention.

CC Sequence 428 AA;

Query Match 36.3%; Score 866.5; DB 19; Length 428;
 Best Local Similarity 42.3%; Pred. No. 5.8e-70;
 Matches 183; Conservative 73; Mismatches 136; Indels 41; Gaps 11;

```

QY 28 LKLNLVRLF--PQKDIYLV-EYSLSTSEFVRN-----YTHVK---DEVREVNCS 74
DB 1 mlnlfrlrlfsytkyfmvlvlsltfsvrlnhkpelfsvshlelgddpnsnvtc 60
QY 75 GIYEOEPLERIGK-SLEI-----RRRIIDLEDDDVYAMTSDCDIYQTLRGYAQLV 124
DB 61 kvlggdpelqkvklellvtqfkkrr-----tphdyimrrdcasftrtkylmepl 114
QY 125 SKEEKSPPIASLVENKDAIMVERLIHAIYQNHNYCYCHYDRKAPDTFKVAMNNLAKCS 184
DB 115 tkeevgfpayslvvhkldmlldrlalypmqpfyylhvdtrkaesflaavgiasced 174
QY 185 NIFIASKLAEVYAHISRLOADLNCISDLKSSIQWKYVNLGQDPFLKSNFELVSELK 244
DB 175 nvfvaqsglesvvaswstrvkdlncmkdyrmanxkylilnlgmdipkltntleivrk 234
QY 245 KLANGANMLETVKPPNSKLERFTYHHELRVPEYVKLPRTINISKEAPPNIQIFVGSAY 304
DB 235 sfgensletekmpnkeerwk-----kryttvvdgkl-tatgyvkaqplktdlfsagay 288
QY 305 FYLSQAQVVKITFNNSIVODEFAMSKDTYSPDEHFWATLIRVGPGEISRSQAQ-DVSDIQ 363
DB 289 fvtvtrvgyvvlennkldkfmewagdclyspdehfwatlgtripevpslpskhydlsdm 348
QY 364 SKTRLVKNMYVEGFF----YPSCTGSHLSVCIYGAELRWLKGHWFRANKFDSKVD 418
DB 349 avarfvkwqfyfegdvnsngapypcsyghvrsvcvfygdsmlrkhhfankfxmdvdp 408
QY 419 ILIKCLAEKLEEQ 431
DB 409 falqlceehlrhk 421

RESULT 7
AAB54344
ID AAB54344 standard; Protein; 465 AA.
XX
AC AAB54344;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:796.
DE
```


CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose germ cells and somatic cells all contain a DNA construct introduced into the animal or an ancestor of the animal at an embryonic stage, where incorporation of the DNA construct into the germ line of the animal causes the animal to develop cardiomyopathy similar to that associated with diabetes mellitus and hyperglycaemia. This sequence represents the human core 2 Gnt protein (beta-1,6-N-acetylglucosaminyltransferase) CC which is used to describe the method of the invention.

XX Sequence 428 AA;

Query Match 36.1%; Score 863.5; DB 19; Length 428;
Best Local Similarity 43.1%; Pred. No. 1,1e-69;
Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

OY 28 LKLLINVRRLF--PQKDIYLV-EYSLSTSPFVNRRTYHKDE---VRY-----EV 71
DB 1 mlrtllrrrlfssypkkyfwmvlslltfsvlr---lhqkpefsvrhlagenpsd1 57
OY 72 NCSGIYEQEPLKIGK-----SLEIRRDITLDDDDVAMTSDCDIYQTLRGYAOXLYS 125
DB 58 nctkviqgdvneikvklleltvkfkrp--rtppdyimtsccsfikrkyivep15 115
OY 126 KEKSPFIAYSLVYHKDAIMVERLTHAIYNOHNYCIHYDRKAPDTEKVMNNLAKCFSN 185
DB 116 keeaeffplaysivvnhk:emldrillraltymqnfycvhtkcsedyslaavmgiascsfn 175
OY 166 IFIAKLEAVEYAHISRLQADLNCSDLLKSSIQWKYVINLCGQDFPLKSNFELYSELK 245
DB 176 vfvasrlsvvyasvsvqadlncmkdlyamsanwylilnlgmdfpiknleivvkl1 235
OY 246 LKGANMLETVKPPNSKLERFTYHHELRVPYEV--KLPIRTNISKEAPPHNIQIVGSA 303
DB 236 lmgennletermpshkeerw-----kryevvngkl-tntgvtvkmplpletplfsgsa 287
OY 304 YFVLSQAFVKYIFNNSTIVQDFPAMSKOTYSPDEHFWATLIRVPGIGELISRSAG-DVSD 362
DB 288 yfvsvreyvgvylqnekikqlmewagdcyspdeylwatigrlpevpaspashkylsdm 347
OY 363 QSKRTLKVMNYYEGF-----YPSCTGSHLSVCIYGAELRWLIDGHPMNFDSKYD 417
DB 348 qavarfvkwgyfegdvskgapypodgyhvrvscifsgaglnmmlrkhhlfankfdvdvd 407
OY 418 PILIKCLAELKEEQ 431
DB 408 lfaicldenhrlhk 421

RESULT 9
AAB30298
ID AAB30298 standard; Protein; 428 AA.

XX AAB30298;
XX 12-FEB-2001 (first entry)
XX Human heart core 2 GlcNAc-T.
XX Human: diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
XX UDP-GlcNAc:galbetal-3galNAcaldiphar betal-6
XX N-acetylglucosaminyltransferase.
XX Homo sapiens.
XX OS
XX US6131578-A.
XX 17-OCT-2000.
XX 02-OCT-1997; 97US-0943058.
XX 02-OCT-1996; 96US-0046876.
XX

PA (KING/) KING G L.
PA (NISHI/) NISHIO Y.
PA (KOYA/) KOYA D.
PA (DENN/) DENNIS J W.
PA (WARR/) WARREN C E.
XX Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
XX WPI: 2000-678642/66.
XX N-PSDB: AAC65469.
XX

PT Preventing or treating cardiomyopathy associated with diabetes mellitus
PT and hyperglycaemia, comprises administering a substance that inhibits
PT core 2 N-acetylglucosamine-T activity -
XX Claim 3; Fig 9; 21pp; English.

CC The present invention describes the human UDP-GlcNAc:galbetal-3galNAc
CC alphaar betal-6-N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
CC and coding sequences. The enzyme is associated with cardiomyopathy in
CC diabetes and hyperglycaemia sufferers. The invention also provides
CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
CC in treatment.

XX Sequence 428 AA;

Query Match 36.1%; Score 863.5; DB 21; Length 428;
Best Local Similarity 43.1%; Pred. No. 1,1e-69;
Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

OY 28 LKLLINVRRLF--PQKDIYLV-EYSLSTSPFVNRRTYHKDE---VRY-----EV 71
DB 1 mlrtllrrrlfssypkkyfwmvlslltfsvlr---lhqkpefsvrhlagenpsd1 57
OY 72 NCSGIYEQEPLKIGK-----SLEIRRDITLDDDDVAMTSDCDIYQTLRGYAOXLYS 125
DB 58 nctkviqgdvneikvklleltvkfkrp--rtppdyimtsccsfikrkyivep15 115
OY 126 KEKSPFIAYSLVYHKDAIMVERLTHAIYNOHNYCIHYDRKAPDTEKVMNNLAKCFSN 185
DB 116 keeaeffplaysivvnhk:emldrillraltymqnfycvhtkcsedyslaavmgiascsfn 175
OY 166 IFIAKLEAVEYAHISRLQADLNCSDLLKSSIQWKYVINLCGQDFPLKSNFELYSELK 245
DB 176 vfvasrlsvvyasvsvqadlncmkdlyamsanwylilnlgmdfpiknleivvkl1 235
OY 246 LKGANMLETVKPPNSKLERFTYHHELRVPYEV--KLPIRTNISKEAPPHNIQIVGSA 303
DB 236 lmgennletermpshkeerw-----kryevvngkl-tntgvtvkmplpletplfsgsa 287
OY 304 YFVLSQAFVKYIFNNSTIVQDFPAMSKOTYSPDEHFWATLIRVPGIGELISRSAG-DVSD 362
DB 288 yfvsvreyvgvylqnekikqlmewagdcyspdeylwatigrlpevpaspashkylsdm 347
OY 363 QSKRTLKVMNYYEGF-----YPSCTGSHLSVCIYGAELRWLIDGHPMNFDSKYD 417
DB 348 qavarfvkwgyfegdvskgapypodgyhvrvscifsgaglnmmlrkhhlfankfdvdvd 407
OY 418 PILIKCLAELKEEQ 431
DB 408 lfaicldenhrlhk 421

RESULT 10
AAB18999
ID AAB18999 standard; Peptide; 406 AA.

XX AAB18999;
XX 08-FEB-2001 (first entry)
XX A core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region.
XX

XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
XW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KM septic shock; adult respiratory distress syndrome; ARDS; cancer;
KV platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
RW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis.
XX
OS Homo sapiens.
XX
CA296936-AI.
PN
XX
03-AUG-2000.
PD
XX
03-FEB-2000; 2000CA-2286936.
PF
XX
PR 03-FEB-1999; 99US-0118674.
PA
XX
(GLXC-) GLICODESIGN INC.
PI
XX
Korczak B, Lew A;
PL
XX
MPI; 2000-594746/57.
DR
XX
New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
PS Claim 3; Page 53; 66pp; English.

The present sequence is derived from a human core 2
beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
The polypeptide can be used to treat diseases and disorders, such as
cancer, cardiovascular disorders and inflammatory disorders including
asthma, rheumatoid arthritis, inflammatory bowel disease,
arteriosclerosis, septic shock, adult respiratory distress syndrome
(ARDS) and cancer. Various platelet-mediated pathologies such as
atherosclerosis and clotting can also be treated. The polypeptides of
the invention are predominantly expressed in gastrointestinal tissue
(stomach, colon, intestine, testis) and are elevated in cancer.
Gastrointestinal disorders that may be prevented or treated include
ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
and ulcerative colitis. The antibodies may be used in
immuno-histochemical analysis, to detect the novel polypeptide and to
localize it to particular cells and tissues and to specific subcellular
locations and to quantitate the level of expression.

Sequence 406 AA:

Query Match 36.1%; Score 862; DB 21; Length 406;
Best Local Similarity 44.5%; Pred. No. 1,4e-69;
Matches 173; Conservative 70; Mismatches 128; Indels 18; Gaps

51 STSPFVRN-RYTHVKDVEHYVNCIGIT--EDEL--EIQSLSEIRRDITLEDODVA.105
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
15 sgqgryrnllymlkikpakrsincsgvttcgdaavqalinnlvkkrr-epftchys 73
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
106 MTSDCDIOFLRBYAOKIVSKKEEKSPINSLVHKKDAIVERLIIAHYNORHTCYHD.165
||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
74 IADTFKHFKETFFIFGISKEEVLEPISAWTIHKIEFTFERLLAVAGPQLIYCVD 133
:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
166 RKAPDFKVVAMNNLAFCFSNIPTASKLEVAYEHISRLQADLCSDLKSSIQMKVIYN 225
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
134 ekspefkvekvaisacfpnvflasklivrvyaswargadlncomedllqsvpwkylfn 193
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
226 LGGDPEPKSNFEULSVLKLGANMLETWKPMNSKEREPTEHHELRRVPYEYVKLPIRT 285
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
194 tctgtcfpkksnaeevgalklmngnrnsesevphkreturkyhfeyvr---dlhl---t 247
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
286 NISKRAPNHNIQTIVSGAIVLVSOAFVKYIFFNMSIYODEFAWSKDITYSDEHWATILRV 345

Db	248	nkkdkppnyltmftgnayivasrdifqghvllknpksqqliewvktxyspdehlwatlgira
Cy	346	PGRPER-SRAQOVSDAQSTRRLVKNMYGCF-----YPSCHSHLSVCIGAEELR
Dd	308	rwmgsypnhkyddtsmtlsarlrvkwgshgdldkgaparpccslngtralcyagadln
Oy	400	WLKDGHWFKNFDSKYDPIILCKLEKL 428
Db	368	wlmgnhllankfdpkvdnalgcleeyl 396
 RESULT 11.		
ID	AAR51386	standard; Protein; 428 AA.
AC	AAR51386;	
DT	07-OCT-1994	(first entry)
XX		
DE	Sequence of human core 2 beta 1-6	
DE	N-acetylglucosaminyltransferase (C2Gnt or core 1-6 AGT).	
XX		
KW	C2GNT: 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;	
KW	O-glycan.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	10..32
FT	Modified-site	/label= signal/membrane anchoring domain 52 /label= potential N-glycosylation site /note= "see also AAs 58 and 95"
PN	EP590747-A.	
PD	06-APR-1994.	
PF	29-SEP-1993:	93EP-0250268.
PR	01-OCT-1992:	92US-0955041.
XX		
PA	(LJOL-) LA JOLLA CANCER RES FOUND.	
PI	Bierhuizen MFA, Fukuda M;	
DR	WPL: 1994-111195/14.	
DR	N-FSDB; AAO61559.	
PT	New beta 1-6-N-acetylglucosaminyl transferase and acceptor - are used for the study of the effect of variant O-glycan(s) on cell-cell interactions, partic. in cancers	
PS	Disclosure: Page 20-22; 34pp; English.	
CC	C2GNT, or an active fragment thereof, catalyses the formation of critical branches in O-glycans. CDNs encoding various glycosyltransferases can be isolated by transient expression of cDNA in recipient cells, e.g. COS-1. COS-1 cells were transfected with a cDNA library, pCDR alpha-PF1, constructed from poly(A)+ RNA of activated T lymphocytes which express the C2Gnt. Transfected cells were selected using Mab T305, which identifies a hexasaccharide on leukostialin. Leukostialin CD43 is an acceptor molecule for C2Gnt activity. A plasmid, pCDR alpha-leu, which directed expression of the T305 antigen was identified. The cDNA insert was isolated and sequenced. The cDNA encoded the acceptor molecule leukostialin CD43.	
XX	Sequence 428 AA:	

Query Match 35.8%; Score 854.5; DB 15; Length 428;

Best Local Similarity 42.9%; Pred. No. 7.1e-69;
Matches 186; Conservative 72; Mismatches 133; Indels 43; Gaps 127

[illegible]

RESULT 12

ID AAB30297 standard; Protein; 427 AA.

AC AAB30297;

DT 12-FEB-2001 (first entry)

DE Diabetic rat heart core 2 GLCNAc-T.

Rat; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia; KW

N-acetylglucosaminyltransferase.

OS Rattus sp.

US6131578-
PN

PD 17-OCT-2000.

02-OCT-1997;
PF

02-OCT-1996; 96US-0046876..

PA (KING/) KING G L.

PA (KONA/) KONA D.

PA	(WARR/)	WARREN C E.
PA	(DENY/)	DENNIS C W.

PI Nishio Y, Koya D,

AA
DR WPI; 2000-678642/6

DR N-FCDD, HANCOCK CO. CT
XX

PT and hyperglycemia, preventing or crea

PT core 2 N-acetylglucosamine-T activity
XX
PS Example 1; Flg 3B; 21pp; English.

The present invention describes

CC The present invention describes the human UDP-glucose:galactose-3galNAc
CC alpha1beta1-6-N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
CC and coding sequences. The enzyme is associated with cardiomyopathy in
CC diabetes and hyperglycaemia sufferers. The invention also provides
CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
CC in treatment.

Sequence 427 AA;

Query Match	Score	DB	Length
35.68;	850;	21;	427;

Matches	178;	Conservative	78;	Mismatches	133;	Indels	46;	Gaps	12,
---------	------	--------------	-----	------------	------	--------	-----	------	-----

[illegible]

RESULT 13

ID AAR71932 standard; Protein; 400 AA.

AC AAR71932

DT 23-SEP-1995 (first entry)

DE I-branching enzyme.

Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IGNT; KW

KW transgenic animal: I antigen.

Homo sapiens.

PN W09507020-A.

PD 16-MAR-1995.

09-SEP-1993; PF

PR 09-SEP-1993; 93WO-US08476.
 XX (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 XX Blerhuizen MFA, Fukuda M;
 PI
 XX MPI: 1995-123182/16.
 DR N-PSDB; AA089201.
 DR
 XX Nucleic acid encoding I-branching enzyme - used to develop prods.
 PT for treating e.g. inflammatory responses, tumours or
 PT hypersensitivity reactions
 PS
 XX Disclosure; Fig.6: 56pp; English.
 CC Poly-A RNA isolated from human PA-1 teratocarcinoma cells was
 CC reverse-transcribed and inserted into expression vector pCDNA1.
 CC Plasmid DNA was used to transfect CHO-Py-leu cells. Transfected
 CC cells were screened using human anti-I antigen antibodies and goat
 CC anti-human Igm. Plasmid DNA was analyzed to obtain DNA (given in
 CC AA089201) encoding I-branching enzyme (AA071932).
 CC
 XX
 SQ Sequence 400 AA;

Query Match 29.8%; Score 711; DB 16; Length 400;
 Best Local Similarity 43.7%; Pred. No. 5.9e-56;
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SDCDIYOTLNGYAOKLVSKKEKSPFIAYSLVVKDAIMVERLLHAIYNOHNYCIHYDRK 167
 DB 70 sscKEYltqshytaplskseeadfpIayImvIhhtftarIfralYmpqnlYcvhvdex 129
 QY 168 APDFFKVMNNLAKCFSENFIFASKLEAVEYAHISRLOADNCLSDLKSSIQKRYVINLC 227
 DB 130 atterfkaveqliscfpnafaskmevpyvgisrlqadlnclrdtsafevskYvintc 189
 QY 228 GODPPLKSNFELVSELKLGANLLEVKPPNSKLEFRTY-HHELRVPEYVKLPRTN 286
 DB 190 gddfpIktkneIvqYlkgfkgkntIpgvIppahalgtrtkYvqehlgkelsYv---Iltt 246
 QY 287 ISKEAPPHNIQIFVGSAYFVLSQAFVKYIFENNSIVODFFAMSKDTPSPDEHFVATLIRVP 346
 DB 247 alkpppphnltyfgsayaIstefanfyIhdpvavdlIqskdtfspehfwvItnirp 306
 QY 347 GIPGEISRSADVSDLOSKTRLVKWNYYEGFFYPSCTGSHRSVCITGAAELRWLIRDOGH 406
 DB 307 gvpqsmnas-----wtgnIraIkwdsmed-rhgchgHyvngIcYngdIkwIvnsps 360
 QY 407 WFANKFDSKYDPIILIKCLAEKLEEQOR 433
 DB 361 lfankfelntYplvecl--elrhrer 385

RESULT 14
 AAR92474
 ID AAR92474 standard; Protein: 400 AA.
 XX
 XX AAR92474;
 AC
 XX 22-APR-1996 (first entry)
 DT
 XX Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme.
 DE
 XX Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme.
 KW glycosyltransferase; blood group; I antigen; poly(lactosaminoglycan.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Domain 7..25
 FT /label= Membrane-anchoring_domain
 FT Modified-site 37

FT Modified-site 212 /label= N-glycosylation_site
 FT Modified-site 255 /label= N-glycosylation_site
 FT Modified-site 255 /label= N-glycosylation_site
 FT Modified-site 314 /label= N-glycosylation_site
 FT Modified-site 388 /label= N-glycosylation_site
 FT
 XX US5484590-A.
 FN
 XX 16-JAN-1996.
 PD
 XX 09-SEP-1993; 93US-0118906.
 PF
 XX 09-SEP-1993; 93US-0118906.
 PR
 XX (LJOL-) LA JOLLA CANCER RES FOUND.
 PA
 XX Blerhuizen MFA, Fukuda M;
 PI
 XX MPI: 1996-087019/09.
 DR N-PSDB; AAT16201.
 DR
 XX Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme
 PT - used to develop prods. for the study, detection and treatment of
 PT pathological conditions involving the enzyme.
 PS
 XX Claim 1; Column 31-34; 29pp; English.
 CC Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme
 CC (Igt) (AAR92474) is an enzyme that converts blood group I antigen to
 CC I antigen. The occurrence of I antigen is associated with
 CC development and maturation of erythroid cells. Membrane-bound or
 CC soluble forms of Igt are obt. by expression of encoding DNA
 CC sequences (see AAT16201) in host cells. Igt is used to study the
 CC role of Igt in development and oncogenesis, to alleviate disorders
 CC associated with Igt underexpression (e.g. haemolytic disease of
 CC the newborn, haemolytic anemia and thrombocytopenia), to regulate
 CC Igt expression, and in detection and diagnostic applications.
 CC
 XX
 SQ Sequence 400 AA;

Query Match 29.8%; Score 711; DB 17; Length 400;
 Best Local Similarity 43.7%; Pred. No. 5.9e-56;
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SDCDIYOTLNGYAOKLVSKKEKSPFIAYSLVVKDAIMVERLLHAIYNOHNYCIHYDRK 167
 DB 70 sscKEYltqshytaplskseeadfpIayImvIhhtftarIfralYmpqnlYcvhvdex 129
 QY 168 APDFFKVMNNLAKCFSENFIFASKLEAVEYAHISRLOADNCLSDLKSSIQKRYVINLC 227
 DB 130 atterfkaveqliscfpnafaskmevpyvgisrlqadlnclrdtsafevskYvintc 189
 QY 228 GODPPLKSNFELVSELKLGANLLEVKPPNSKLEFRTY-HHELRVPEYVKLPRTN 286
 DB 190 gddfpIktkneIvqYlkgfkgkntIpgvIppahalgtrtkYvqehlgkelsYv---Iltt 246
 QY 287 ISKEAPPHNIQIFVGSAYFVLSQAFVKYIFENNSIVODFFAMSKDTPSPDEHFVATLIRVP 346
 DB 247 alkpppphnltyfgsayaIstefanfyIhdpvavdlIqskdtfspehfwvItnirp 306
 QY 347 GIPGEISRSADVSDLOSKTRLVKWNYYEGFFYPSCTGSHRSVCITGAAELRWLIRDOGH 406
 DB 307 gvpqsmnas-----wtgnIraIkwdsmed-rhgchgHyvngIcYngdIkwIvnsps 360
 QY 407 WFANKFDSKYDPIILIKCLAEKLEEQOR 433
 DB 361 lfankfelntYplvecl--elrhrer 385

Fri Aug 3 09:38:08 2001

us-09-645-392-2.rag

Page 12

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: August 2, 2001, 14:43:34 ; Search time 25.93 Seconds
(without alignments)
1330.778 Million cell updates/sec

Title: US-09-645-1.92-2

Perfect score: 2389
Sequence: 1 MKRIFKCYFHHTQOKYFILE.....DWITLPSEKLFMRRLTTTS 453

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	863.5	36.1	428	A46293	beta-1,3-galactosyl
2	711	29.8	400	A46297	beta-1,6-N-acetylgl
3	326	13.6	445	T21262	hypothetical prote
4	323	13.5	425	T20207	hypothetical prote
5	315.5	13.2	454	T21261	hypothetical prote
6	296	12.4	401	T24929	hypothetical prote
7	294	12.3	472	T21288	hypothetical prote
8	275	11.5	402	T24930	hypothetical prote
9	245	10.3	459	T24742	hypothetical prote
10	229.5	9.6	454	T25382	hypothetical prote
11	220.5	9.2	489	T24013	hypothetical prote
12	219.5	9.2	470	T23137	hypothetical prote
13	219.5	8.5	486	T21566	hypothetical prote
14	203.5	8.2	472	T23384	hypothetical prote
15	202.5	8.5	753	T24745	hypothetical prote
16	197.5	8.3	448	D85042	probable glycosyl
17	193.5	8.1	478	T27714	hypothetical prote
18	188	7.9	467	T21408	hypothetical prote
19	187	7.8	447	T20096	hypothetical prote
20	182.5	7.6	513	T20068	hypothetical prote
21	178	7.5	367	T28892	hypothetical prote
22	175.5	7.3	406	F96571	hypothetical prote
23	168	7.0	395	B96735	unknown protein P2
24	166.5	7.0	434	T51450	probable RING zinc
25	132	5.5	630	T02524	hypothetical prote
26	131	5.3	384	T08940	hypothetical prote
27	126	5.3	371	T21819	hypothetical prote
28	116	4.9	895	E64431	UDPglucose 6-dehyd
29	108.5	4.5	661	S50734	WH13 protein - yea

30	108	4.5	702	2	E69498	hypothetical prote
31	107	4.5	717	2	S78177	hypothetical prote
32	107	4.5	1072	2	T37742	serine threonine-p
33	107	4.5	2265	2	T26193	hypothetical prote
34	106.5	4.5	1347	2	T41321	BTB domain and Ank
35	104	4.4	364	2	T48628	hypothetical prote
36	104	4.4	2485	1	H71621	serine/threonine-s
37	104	4.4	2628	2	S59413	probable membrane
38	103.5	4.3	588	2	H82877	hypothetical prote
39	102.5	4.3	415	2	I65223	heparin-binding fi
40	102.5	4.3	707	2	A54846	fibroblast growth
41	102.5	4.3	707	2	A38429	keratinocyte growt
42	102.5	4.3	769	2	S16236	fibroblast growth
43	102.5	4.3	820	2	S17295	fibroblast growth
44	102.5	4.3	821	1	TVMSBK	fibroblast growth
45	102.5	4.3	821	1	TVHUF2	fibroblast growth

ALIGNMENTS

RESULT 1

A46293
beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
R:Accession: A46293
R:Biernutzen, M.F.; Fukuda, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 9326-9330, 1992
A:Title: Expression cloning of a cDNA encoding UDP-GlcNAc:gal beta 1-3-galNAc-R (GlcN
A:Reference number: A46293; PMID:93028457
A:Accession: A46293
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-428 <BIEM>
A:Cross-references: GB:97347; NID:9183440; PIDN:AAA35919.1; PID:9183441
A:Note: sequence extracted from NCBI backbone (NCBI:115900, NCBI:115901)
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	36.1%	Score	863.5;	DB 2;	Length	428;			
Best Local Similarity	43.1%	Pred	No. 6.2e-56;						
Matches	187;	Conservative	72;	Mismatches	132;	Indels	43;	Gaps	12
QY	28	LLKLLNRRRL--PQDIYLV-EYSISTSPVRRKTHVDE--VRY-----EV 71							
DB	1	MLRLLRRRLFSYPTKYFMVLVSLTFVYLR--IHQPEVSVHLELAGNPSSDI 57							
QY	72	NCSGIYEOEPLDIGK-----SLEIRRDITLDEDDVAMTSDCDIYQTLRGYAKLV 125							
DB	58	NCTVLOGDVNEIQVKKEILTVKFKRP--RMTPDYINMTSDCSSFIKRRKIYVPLS 115							
QY	126	KEESFPLAYLVVYHKDAIMVERLIHAIYNOHNTYCHYDRKADTKVAMNNAKQFSN 185							
DB	116	KEEFEPFAYSIYVHHKTEMLDRLRALYQPNQYCHVDKSSDSYLAAMGJASQFSN 175							
QY	186	IFIAKLEAVYAHISRLQADLNCISDLKSSIQMKYVINLGGDFPKSFTFELVSELK 245							
DB	176	VFVASRLSESVYYASMSRQADLNCMKDLYAMSAAMKYLINCQMDPIKTLTEIVRLKL 235							
QY	246	LNGANMLETVKRPNSKLERFTYHHELRVPEYV--KLPIITNSKAPPNINIOIFGSA 303							
DB	236	LMGEHNLETERMPHSKHEERW-----KKRAYVNGKL-TNTGTVKMLPLETPELFSGSA 287							
QY	304	YFVLSQAFVYKTIFFNNSIYQDFFAWSKDTYSPDEHFMAILLVPGIPGEISASAO-DVSDL 362							
DB	288	YFVVSREIVGYLVQNEKIOKLMEANQDTPSPDEYIMATIQIIPVPGSLPASHKYDLSDM 347							
QY	363	QSKTRLVKWNYYEGFF-----YPSCTGSHLSVSCIYGAELRMILTKDGMFANFKFSKVD 417							
DB	348	QAVARFVVKQVFEQDVSGAGAPYPCDGVHVSVCIFGAGDLNMMILRKHLFANKEEDVDV 407							

QY 418 PILIKCLAEKLEEQ 431
 Db 408 LFAICLDEHLRHK 421

RESULT 2

A46297
 beta-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.-) - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A46297
 R:Blotting: M.F.; Mattei, M.G.; Fukuda, M.
 Genes Dev. 7, 468-478, 1993
 A:Title: Expression of the developmental I antigen by a cloned human cDNA encoding a mem
 A:Reference number: A46297, MID:93194065
 A:Accession: A46297
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-400 <RLE>
 A:Cross-references: GB:219550; NID:q296531; PIDN:CAA79610.1; PID:q296532
 A:Note: sequence extracted from NCBI backbone (NCBIN:127166, NCBI:127167)
 C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 29.8%; Score 711; DB 2; Length 400;
 Best Local Similarity 43.7%; Pred. No. 9e-45;
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SDCIYOTLGYAOKLSEKSPFIAYSLVHKDAIMVERLIHAIYNQHNICYHDKR 167
 Db 70 SSCSEYLTQSHYITAPLSKEADDFLAYIMVHHHPTFARLFRALYMPQNIICVHVDK 129
 QY 168 APDTFKYANNLAKCFENITASKLEAVEYAHISRLADINCLSDIKSIOKKYIINC 227
 Db 130 ATTEFKAVDQLSCFPNAPLAKMEPVYVGGISRLQADNCRDLSAFESWKYIINC 189
 QY 228 GODEPLKSNFELKLGKNGANLKVPPNSKLERFTY-HHELRYVEYVKLPRTN 286
 Db 190 GODEPLKTNKEIYVGLGFKGNITTPGLPAHIGTKRVHGHJGKELSYV---IRTT 246
 QY 287 ISKRAPPHNIOIEVGSAYFVLSQAFVYKIFNNSTIVODFPMSKDTYSPDEHFAATLRVP 346
 Db 247 ALKPPPHNLTIVYGSAYVALSRFANFVHDPRAVDLLQMSKTFSPDHFVTLNRIP 306
 QY 347 GIPEISRSADVDLSQKTLVKNVYEGFPYSGTSHLSVCIYGAELMLIKDGH 406
 Db 307 GVPSSMKNAS-----WTGNLRAIKMSDMED-RHGGCHGHVHGCIYGNODLKLNVSPS 360
 QY 407 WFANKFSDVPIILIKLAEKLEEQOR 433
 Db 361 LFANKFELNTPVTECL-ELRHRER 385

RESULT 3

T21262
 hypothetical protein F22D6.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21262
 R:Milkinson, J.
 submitted to the EMBL Data Library, April 1996
 A:Reference number: 219397
 A:Accession: T21262
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-445 <WIL>
 A:Cross-references: EMBL:T21262; PIDN:CAA95817.1; GSPDB:GN00019; CESP:F22D6.11
 A:Experimental source: clone F22D6
 C:Genetics:
 A:Gene: CESP:F22D6.11
 A:Map position: 1
 A:introns: 65/3; 95/2; 136/3; 180/3; 256/3; 281/3; 403/3

Query Match 13.6%; Score 326; DB 2; Length 445;
 Best Local Similarity 25.1%; Pred. No. 2e-16;
 Matches 108; Conservative 85; Mismatches 176; Indels 62; Gaps 15;

QY 25 LSLKLNVRLRLEPPQKDIYLVESLSTSPFVRRNRYHVKDEV--RYVNCSTIY----E 78
 Db 17 LVIFLYLFIVPLVDR--ILESLNSRNPQETSLSKIEHDLDDIDINDINLFNGSKN 73
 QY 79 QEPLEIGSLERRDDIILEDVVAMTSPCDIYQITRGYAKLSEKSPFIAYSLV 138
 Db 74 RQQLIVNSRST-----EDKILYSTDKQYTLKSLFRNKPKPLPEESFPLSLGL 124
 QY 139 VKRD--AIMVER--LIHAIYNQHNICYHDKRAPDTFKYANNLAKCFENITASKLE 193
 Db 125 YKELSQVTIQVLEPMLTSIYHPCNECYIAVENSAPIFQNLKELSCFSNIHF-MKRP 183
 QY 194 AVEYAHISRLADINCLSDIKSIOKKYIINC GODEPLKSNFELKLGKNGANL 253
 Db 184 PIDWOSHETINSAYDCLFSLHLSKDMRYFOYLSGVDLPLTNLEMQIILKHLNG----- 238
 QY 254 TVKPPNSKLERFTYHHELRYVEYVKLPRTNISKRAPPHNIOIEVGSAYFVLSQAFVK 313
 Db 239 -----TANVEIK--PYQYQL--RGNKTQSP--LPLFKSLISLIPREAN 279
 QY 314 YFPNNSIYODFPMSKDTYSPDEHFAATLI---RVPGIPEGISASQVSDLSQKTRLV- 369
 Db 280 HLSSSIPQQLLEFLRNGLADEGFWGLFGNKNLFTDPSGLNKRKEMISYNNVETLTLY 339
 QY 370 ---KNVY---EGFPYSGTSHLSVCIYGAELMLIKDGHFANKFSDVPIILIK 422
 Db 340 PIDGMRYYISRQIYISKRNCHYMKAGSCVSGIGVPLLSKALVAIKFKLSPEAYF 399
 QY 423 CLAEKLEEQOR 433
 Db 400 CL---LKEHRR 407

RESULT 4

T20207
 hypothetical protein C54C8.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T20207
 R:Doobson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: 219237
 A:Accession: T20207
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-425 <WIL>
 A:Cross-references: EMBL:T20207; PIDN:CA805469.1; GSPDB:GN00019; CESP:C54C8.11
 A:Experimental source: clone C54C8
 R:McLay, K.
 submitted to the EMBL Data Library, October 1998
 A:Reference number: 220372
 A:Accession: T20372
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-425 <WIL>
 A:Cross-references: EMBL:AL032664; PIDN:CAA21769.1; GSPDB:GN00019; CESP:C54C8.11
 A:Experimental source: clone Y91F4A
 C:Genetics:
 A:Gene: CESP:C54C8.11
 A:Map position: 1
 A:introns: 29/2; 117/3; 161/3; 195/3; 237/3; 266/3; 295/1; 313/3; 355/1; 374/2

Query Match 13.5%; Score 323; DB 2; Length 425;
 Best Local Similarity 22.6%; Pred. No. 3.1e-16;
 Matches 102; Conservative 85; Mismatches 176; Indels 88; Gaps 14;
 QY 16 VFILFLMLLSILKLLNVRLE-PQKDIYLVESLSTSPFVRRNRYHVKDEVRYVNC 74

```

Db 5 1FLPFLIEFTFLVLEIFESQNPASTIDPS-----HECQ 44
QY 75 GYEOEPLEIGKSLERRRDIIDLEDDVYAM-----TSDCIYQTLRGYAQKLVSK 126
Db 45 KMFQGDVSLA-----RGALFKFDREIILAQILKLSGEOACAEFQKJFGFQEPSTQ 98
QY 127 EKEFPFLAIVKADIMVERLHAITYNOHNYCHIDRKAPDTEFVAMNNLAKCSNI 186
Db 99 EELFEPFLAYGLVHGDVQSLSLATYQPNQFCIAVDGNSVEFTGLVPMISRCIGN 158
QY 187 --FLASKLEAVEYAHISRLQADINCLSDLLKSSIQWKYVNLGQDFPLKSNFELVELK 244
Db 159 QYFILDRLMCGYELLTSV---FOCVYLAFLPSDMKYFOYLSGVADAPLNMELMIRLK 215
QY 245 KLNGANMLEYKPPNSKLERETTYHEELRRVPEYEVKPIRTNISKEAPPHNIQFVSAY 304
Db 216 ALNG-----SFNAEI--LPFEFYRLNRKRPWSSPLPKTSL--SAT 253
QY 305 FVLSQAFKYIFNNSIVODFPAMSKDTYSPDEHFMTLIRVPGI-----PGEISRSA 356
Db 254 F--SRKANFVNSEKYLEQIDFLRGTTCADESIMATIAQPKYIKRSELPMBGFDPAKA 311
QY 357 QDVSDL-QSKTRLYKM-----NYEGFPY-----SCGSHLRSCYIGAAELRW 400
Db 312 WIKNKYRTRGKLGKYGEMOKIDNGYVSRYOQYVNRAPVKCKGYRRLSCVFGYDLPN 371
QY 401 LKDGHWPFANKFDSKVPDLILKLAELKEQ 431
Db 372 LINRHELVAKHLYSYQPAAMCIVENSROK 402

```

RESULT 5

```

T21261
hypothetical protein F22D6.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21261
R:Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: 219397
A:Accession: T21261
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-454 <WIL>
A:Cross-references: EMBL:Z71262; PIDN:CAA95816.1; GSPDB:GN00019; CESP:F22D6.12
A:Experimental source: clone F22D6
C:Genetics:
A:Gene: CESP:F22D6.12
A:Map position: 1
A:Introns: 62/3; 92/2; 190/3; 266/3; 291/3; 412/3

```

Query Match 13.2%; Score 315.5; DB 2; Length 454;

Best Local Similarity 24.3%; Pred. NO. 1.2e-15;

Matches 111; Conservative 79; Mismatches 190; Indels 77; Gaps 16;

```

QY 8 EKHITLQAKYFLFLMLLSLKLNLNVRLLFPQKDIYLVESISGPSVRNRHYTHKDEV 67
Db 6 FETSLIAIFLFI-----YFSVESLFRKO-----EDKNVSKOLKSLCTIASDSY 52
QY 68 ---REVNCGIYEQEPLEIGKSLERRRDIIDLEDDVYAMTSDCIYQTLRGYAQKLV 124
Db 53 LADNNEINCSNII--KGYKTNEKLDIMHDI--EOLFSCQTKCOTLKTFLFRNTNPM 107
QY 125 SKEEKSFPFLAYSLVYHKDA-----IMVERLHAITYNOHNYCHIDRKAPDTEFV 166
Db 108 SAEEKFEPFLSYGMLYKKDLPQVTPARMTKHLFLEINLOVLEFLSSITHPQNEYIANG 167
QY 167 KAPDPFKVAMNNLAKCFNSIFLASKLEAVEYAHISRLQADINCLSDLLKSSIQWKYVNL 226
Db 168 NSAPFIFQNLREVSTCFSVVHF-MKRPPISWGSHEIIDSVDCLFELSLHLETDWRYFOYL 226

```

```

QY 227 CGDPEPLKSNFELVSELKKNLNGANMLEYKPPNSKLERETTYHEELRRVPEYEVKLPRTN 286
Db 227 SGVDIPLKTNLEKVOYLLKLNSTNVEIT---NYQARLTGNE----- 267
QY 287 ISKEAPHNIQIFVGSAYVLSQAFKYIFNNSIVODFPAMSKDTYSPDEHFMTLIRVPG 346
Db 268 --NESP---LPLFKSLSAIIPRKAANQLASNTARKLLEFLMNTETLADGCFWGTLPGNK 322
QY 347 ---GIGET-SRQADVSDLSQK--TRLVKMNY---EGEFYPSCTGSHLR--SYCIYGA 396
Db 323 DQFNISGSINSDKMYRDNQNNIENPTDGMSTYISRDIMPELCKMYKDKSCVFGIG 382
QY 397 ELRFLIKDGHWPFANKFDSKVPDLILKLAELKEQ 433
Db 383 DVPRLRTSKALVAHKFYLKSEPEAYFCL--LKEHHR 416

```

RESULT 6

```

T24929
hypothetical protein T15D6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24929
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: 219956
A:Accession: T24929
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-401 <WIL>
A:Cross-references: EMBL:283125; PIDN:CAB05620.1; GSPDB:GN00019; CESP:T15D6.2
A:Experimental source: clone T15D6
C:Genetics:
A:Gene: CESP:T15D6.2
A:Map position: 1
A:Introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3

```

Query Match 12.4%; Score 296; DB 2; Length 401;

Best Local Similarity 23.4%; Pred. No. 2.8e-14;

Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;

```

QY 61 THYKDEVREYVNCSGIYEQEPLEIGKSLERRRDIIDLEDDVYAMTSDCIYQTLRGYA 120
Db 34 TDLIDL--ELNCTNII--QGLNKEQLELINTKMT---EDKLMNSTDRCHTVKSMFRFN 86
QY 121 QKLVSEKSEKFPFLAYSLVHKDAILMERLHAITYNOHNYCHIDRKAPDTEFVAMNNLA 180
Db 87 EYPLSEEEARFPPLSYGLVRELQVLFMLSTIYQPNQEXCIAGVENSASTFILLLELS 146
QY 181 KCFGNIFASKLEAVEYAHISRLQADINCLSDLLKSSIQWKYVNLGQDFPLKSNFELY 240
Db 147 DCFPN-----KRPITMGSEIINSYVDCLKFSLKSMNKRYQYLSGVDPILKTMLEMY 201
QY 241 SELKKNLNGANMLEYKPPNSKLERETTYHEELRRVPEYEVKLPRTNISKEAPHNIQIY 300
Db 202 RILKSLNGANVETIKYENRRL-----LGQNETESPPLFK 237
QY 301 GSAFYLSQAFKYIFNNSIVODFPAMSKDTYSPDEHFMTLIRVPG---IPGEISRSAQ 357
Db 238 SLSLSLIPRKAANYLAASSIPQOLLEPLRNTVWADBEFWGTLLGNGGLFDPVPSLN---- 293
QY 358 DVSDLSQKTRLVKNWYEGFFYPSCTGSHLRVSCYIGAAELRWLKDGHWFANKFDSKYD 417
Db 294 -----FEEHQIFESGCHNMKRDGSCVFQIGVSNLLOKALVAHKLVLTS 340
QY 418 PILIKCL 424
Db 341 PEAYFCL 347

```

RESULT 7

T22188

hypothetical protein F44F4.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22188

R:Colles: L.

submitted to the EMBL Data Library, September 1994

A:Reference number: Z19528

A:Accession: T22188

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-472 <WIL>

A:Cross-references: EMBL:Z37092; PIDN:CA85457.1; GSPDB:GN00020; CESP:F44F4.6

A:Experimental source: clone F44F4

A:Gene: CESP:F44F4.6

A:Map position: 2

A:Introns: 57/2; 105/2; 185/3; 261/3; 319/1; 370/1; 412/3

Query Match 12.3%; Score 294; DB 2; Length 472;

Best Local Similarity 21.8%; Pred. No. 4.9e-14;

Matches 103; Conservative 83; Mismatches 184; Indels 102; Gaps 17;

32 LNVR-----LPQKDIYL-----VEYSLSPFVRNRHYHVKDEVREYVNCSGIYE 78

24 MNVRFKTIQSLFYSFTYIVLRLGFDYSL-----VPENRK-----LE 63

79 QEPLEIGKSLF--IRNDI-----IDIEDDVVAMSDCDIYQTLGVAQKLVSK 127

64 NIPLECNVINGTENRRKISRAROMDMHFDVVEHEIFNSRVCSITDKYFNFTIRPSSRE 133

128 EKSPPIAVSLVHDAIVVERLIHAIYNQNIYCIYHDKRAPDFFKVAAMNLKCFN-- 185

124 EAEPLAIGLVYTYIYQVLTQMSLFYQPMFCITVDOSPNRYKRYVIALPSCFFNMH 183

186 IFASKLEAVEYAHISRLQADLNLCLSDLLKSSIQWKYVINLCGDFPLKSNFELSELK 245

184 VFIG---EPSQMGSGFGILKNVYTCFNMLSKSKQMKYQYLSGTDLPIRTNLEWVRLEFKA 240

246 LKGANMLETVKPPNSKLERFTYHHELRVPEYVKLPRTNISKAPPHNIOIFVGSAYF 305

241 LKNGS-----AMTDVSTF-----EVDRY-----KMEGVLEPP--MPYKSSMSV 276

306 VLQSAFYKIFNNSIVQDFAMSKDYSPDEHFNATLIRVPG---IPGEISRSQADVSDL 362

277 VVPEGADYLISPRVQKLKYLKSTWIPDESFMSIVLGSFALLPVPGSI--RVRDILML 334

363 QSKTRL-----VKMNYEGF---FYPSTGSHLSVCITYGAELRWLIKDGHW 408

335 KKNKRLRPYENTYNSITSTIGRYQVWGMQKCEFGKVKDFSCYGVEDIEIETRELV 394

409 ANKEDSKVDPIILKCLAEKLEEO-----RDWITLPSEKLFMDRNL 450

395 AHKILYEPQAFAMCKLEKVRRLSPDAHLFSARSYSQMTVELYOGKAIT 446

RESULT 8

T24930

hypothetical protein T15D6.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T24930

R:Robson, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19956

A:Accession: T24930

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-402 <WIL>

A:Cross-references: EMBL:Z83125; PIDN:CA805621.1; GSPDB:GN00019; CESP:T15D6.3

A:Experimental source: clone T15D6

A:Gene: CESP:T15D6.3

A:Map position: 1

A:Introns: 55/3; 85/2; 121/3; 161/3; 237/3; 262/3

Query Match 11.5%; Score 275; DB 2; Length 402;

Best Local Similarity 24.0%; Pred. No. 9.7e-13;

Matches 101; Conservative 69; Mismatches 169; Indels 82; Gaps 14;

16 VFILFTLMLSLKLLNVRRLFPOKDIYLYEYSTSPFVRNRHYHVKDEVREYVNCSG 75

19 IFILFIKALIRLLENPNQRIRHPTDLL-----DDL--QINCTN 56

76 IYQEPLEIGKSLFIRRDIIIDEDDVVAMSDCDIYQTLGVAQKLVSKNFFIAY 135

57 ILQGFKNF--NLEELINRMT--ENKMNNSDRCQTLTSMRFKVPSEARFPLSF 111

136 SLVHKDAIMVERLIHAIYNQNIYCIYHDKRAPDFFKVAAMNLKCFNFIASKLEAV 195

112 GLIYKELSQVLFLLSSITQPNQENCIANGENSARAFILLLKELANCFN-----KRPPI 166

196 EYAHISRLQADLNLCLSDLLKSSIQWKYVINLCGDFPLKSNFELSELKLGANMLETY 255

167 KNGSEFLINSVYCLEFSLHLSKMDWKYQYLSGVDIPLKTNLEWVRILKRLNG---TV 221

256 KPNKSLERFTYHHELRVPEYVKLPRTNISKAPPHNIOIFVGSAYFVLQSAFYVI 315

222 NIGIS-----TYEDRL-----LNGKNKTESP--LPFKSSLSLIPRAANYL 262

316 FNNISIVQDFAMSKDYSPDEHFNAT-----LINVPPIGELISQADVSDLSKTRLY 369

263 SSSSVYQQLLEFLRTTWVADEGFNGLFGNKDLFVYVPGSFNENDELITNGMGVYSRHL 321

370 KM-----NYEGGFYFSCSTGSHLSVCITYGAELRWLIKDGHWANKFDSKVDPIILK 423

322 -WVSECHNYMD-----NS-CYFGIGDVPNLKSAVAHLIYSESEARFPC 368

424 L 424

369 L 369

RESULT 9

T24742

hypothetical protein T09E11.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T24742

R:McClay, K.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19930

A:Accession: T24742

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-459 <WIL>

A:Cross-references: EMBL:Z81147; PIDN:CA803536.1; GSPDB:GN00019; CESP:T09E11.9

A:Experimental source: clone T09E11

A:Gene: CESP:T09E11.9

A:Map position: 1

A:Introns: 39/1; 78/3; 204/1; 291/1; 338/3; 371/3; 417/3

C:Superfamily: Caenorhabditis elegans hypothetical protein H4IC03.3

Query Match 10.3%; Score 245; DB 2; Length 459;

Best Local Similarity 22.8%; Pred. No. 1.9e-10;

Matches 106; Conservative 74; Mismatches 202; Indels 82; Gaps 17;

43 IYLVEXSLTSPFVRNRHYHVKDEVREYVNCSGIY--EQEPLEIGKSLFIRRDITLED 100

25 VYIYMFSS-SKSLFGRN-----EONSVALYMLNENQORALITNELARDFAKIES 72

101 -----DDVYAMTSDCDIYQTLGVAQKLVSKNFFIAY 132

Db 73 KIEHFSKYRRRSELAHVDCGRILSGDKDYLTQVSG--ENRIPVEMPNDSCSAVMD 130
 QY 133 -----IAVSLVHDKAIVERRLHAIVNQHNYICIHDRKAPDEFFKAA 177
 Db 131 RIITHDHLRLKGVAFARIVYDYELVEKQVMSYHPQSPFCFALDKKAPTFKKQMR 190
 QY 178 NLACFSNFIASKLEAVEVAHISRLQADLNCSDLSKSIOQKYYVNLGQDFPLSNF 237
 Db 191 AMAACLPRVLLLPDQEPIDSGHNVNLAHLNCLRALINKP--GNNYAMLONHMDLLTTSVY 249
 QY 238 ELVSLKTLKNGANMLETQPNKSLERTYHEHLRRVYEVKLPITRNISKEAPPHNIO 297
 Db 250 ELEOVYEWLGGANDVELL--PEAQRLDEENFKMD----PRSLKMPDESKVDETILNKKIK 304
 QY 298 IFVGSAYFVLGSAFYKYIFNNISIVDFE--ANSKDTSPDEHFMATL--IRVPGIPGELSR 354
 Db 305 FSKGVGQSMSPRAAVDMKTRKYNLSTYIDQNNQGMVDEMILISLQISAFGLMPGHF-- 362
 QY 355 SAGQVSDLOSRTLVKMYNVEGFEYPSCTGSHLR--SVCIYGAELRLIKDGHMFANKFD 413
 Db 363 TDQCKEGEPPRMVQWMEDESY---CASKYKRNNVCILGLEYLRSAVSAPFTLMFNKML 419
 QY 414 SKYDEPILIKLAELKEQ---QRDWITLPSEKLFMDRNLTTTS 453
 Db 420 PFDNSITICTAELLNRTFMGQNDH---PLEEYY--KNMWTVS 459

RESULT 10

T25382
 hypothetical protein T27F6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T25382
 R:Dooson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z20026
 A:Accession: T25382
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-454 <MIL>
 A:Cross-references: EMBL:Z82060; PIDN:CAB04881.1; GSPDB:GN00019; CESP:T27F6.1
 A:Experimental source: clone T27F6
 A:Genetics:
 A:Gene: CESP:T27F6.1
 A:Map position: 1
 A:Insertions: 80/3; 191/1; 278/1; 325/3; 365/2; 407/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein H4IC03.3

Query Match 9.6%; Score 229.5; DB 2; Length 454;
 Best Local Similarity 22.7%; Pred. No. 2.5e-09;
 Matches 103; Conservative 68; Mismatches 202; Indels 81; Gaps 19;

QY 17 FIFLFLMLSLKLINVRRLFPQKDIYEVSLSTSPVRRNRYTHVKDEVREYVNCSGI 76
 Db 10 FIL-----LVALICTLN-----WFLYNGV-----YRNHYWGYTONSSPOEDIAKF 50
 QY 77 YEDEPLEIGSLERRDDIIDLDDV---VAMTSCDIQYQLRG-YAOKLYVEKEKSP 132
 Db 51 YPTSNKDY---FYRRRP---ETENYNGQVLAGDTAYLKYTGERTYIAENESLANS 103
 QY 133 -----IAVSLVHDKAIVERRLHAIVNQHNYICIHDRKAPDT 171
 Db 104 CEAVMDKILSRDHLRLGVAFARIVYDYELIEKHVMSYHPQSPFCFALDKKAAKE 163
 QY 172 EKVAANNLACFSNFI-----ASKLEAVEVAHISRLQADLNCSDLSKSIOQK 222
 Db 164 EFERMQAMASCLPRVLLLPGRFFKNPIHDLSDVSHGHNIL--AHYNCLRALINKP--GWN 221
 QY 223 VINCGQDFPLKSNFELVSELKLNGLANMLETQPNKSLERTYHEHLRRVYEVKLP 282
 Db 222 AILLQNHDLITKSYLEKIFNWLGGANDV-AIRPELGRDK-----KHFKMDPMSLK 273

QY 283 IRTNISKAP-----PHNIOIFVGSAYFVLGSAFYKYIFNNISIVDFE--ANSKDTSPDEH 338
 Db 274 LEFNESEIDPVLINTLTKFAKAGAVOSSLSRAAVDMKTRVDTLTIDQNNHGTIVDDQF 333
 QY 339 WAT--LIRVPGIGEIS--RSADVSDLOSRTLVKMYNVEGFEYPSCTGSHLR--SVCIY 394
 Db 334 TQAFQISDFLGMGHFTDCKIKGILTEGITRPAQTTHD---QSKCAKSKSRHGICIMG 390
 QY 395 AAEFLRWLIKDGHWFAKPFDSKVPILIKLAEL 428
 Db 391 IEHLSMAKSEHLMFNKVLPLFDYSITICTAELL 424

RESULT 11

T24013
 hypothetical protein R07B7.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24013
 R:Harris, B.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19830
 A:Accession: T24013
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-489 <MIL>
 A:Cross-references: EMBL:Z75955; PIDN:CAB00115.1; GSPDB:GN00023; CESP:R07B7.6
 A:Experimental source: clone R07B7
 A:Genetics:
 A:Gene: CESP:R07B7.6
 A:Map position: 5
 A:Insertions: 42/3; 80/2; 235/3; 287/1; 372/2; 416/1

Query Match 9.2%; Score 220.5; DB 2; Length 489;
 Best Local Similarity 22.0%; Pred. No. 1.3e-08;
 Matches 102; Conservative 77; Mismatches 209; Indels 75; Gaps 19;

QY 22 TLWLISLKLINVRRLP--QKDIYEVSL-----STSPVRRNRYTH----- 62
 Db 16 TILKISEMSATRYKRLDIDLDADINTARVSLQNGKNSPLIKTRSOEKPRKNTTKMD 75
 QY 63 ---YDEVREYVNCSGIYECEPLEIGKSLERDDIID-----LEDDVYAMTSDC 110
 Db 76 NYFTDDESGF-LNGSEMK-----NNKDVIELYVNNGRMKLDNERLEPLMDC 122
 QY 111 -DIYQTLRGYQOKLYSKEEKSPPIAYSLVHKDALMVERLHAIVNQHNYICIHDRKAP 169
 Db 123 PSIKNRIG---DMSFRLKRPPIAFVRTYIKYELQALISYHDNVFCEYMSKST 179
 QY 170 DFFKAMNNLAKCFENIFASKLEAVEVAHISRLQADLNCSDLSKSIOQKYYVNLGQ 229
 Db 180 DLKKSVAIMSCFTNYYVVLGKEYSLNSGGQDPAHDCDKTLIDR--KMDHAILQNF 237
 QY 230 DPLKSNFELVSELKLNGLANMLETQPNKSLERTYHEHLR--VPEYVKKLPITFNI 287
 Db 238 DLIRTPYQSLDISSELYNTSIM-----GDFGFSYRYTKAKMPAGKLFRIETGV 290
 QY 288 SKEAPPHNIOIFVGSAYFVLGSAFYKYIF---NNSIVDFPAMSKDTSPDEHFMATL 343
 Db 291 PNEILNRLIYKSLNIEYIVSKYFKSMFKLMNIIILKLD-DMDYGVDEMLVQTLYE 349
 QY 344 RVPGIPGEL--RSADVSDLOSRTLVKMYNVEGFEYPSCTGSHLR--SVCIYGAELRW 400
 Db 350 NYLLEGQWESNCTRNHNDI--LTHMTWMDSGPFGFKCKSKWRKRGICIMGEYVNE 407
 QY 401 LIKDGHWFAKPFDSKVPILIKLAEL-----EEQQRDWIT 437
 Db 408 LKISQGVYANKVMAIFDGETIACMEKMKRNTAGETPNTQWLT 450

RESULT 12

T32137

hypothetical protein R07C3.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T32137
 R:Lamar, B.; Kramer, J.
 Submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid R07C3.
 A:Reference number: 221125
 A:Accession: T32137
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1470 <L>AM>
 A:Cross-references: EMBL:AF016686; PIDN:AA566233.1; GSPDB:GN00020; CESP:R07C3.3
 A:Experimental source: strain Bristol N2; clone R07C3
 C:Genetics:
 A:Gene: CESP:R07C3.3
 A:Map position: 2
 A:Introns: 18/1: 78/3; 189/1; 276/1; 319/3; 359/2; 401/3; 441/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein H4IC03.3

Query Match 9.2%; Score 219.5; DB 2; Length 470;
 Best Local Similarity 21.4%; Pred. No. 1.4e-08;
 Matches 95; Conservative 83; Mismatches 207; Indels 59; Gaps 16;

QY 38 FPOKDIYLVEYSLSTSPFVRNRTYHKDEVREYVNSGTYEOP-----LEIGKS 87
 DB 19 FSRNSSLDPLSLVDRRLAYKRTTEALRTYDSCSEOGFYCKRPETOHVDCGRLVGDK 78
 QY 88 LEIRRDIT--LEDDDVVAMTSDCDIYQTLRGYAOKLVEKESKSPFI-----AYSLVH 140
 DB 79 IQLAFQYVQLSLGSRNRIPIENSFLNLCSAISKRIIPKTAQFKRLKNGTAFARIVE 138
 QY 141 KDAIWERLTAIYNQNIYCIHYDKAPDTFKYAMNNLAKCSNFIATSKLAVEYAH 200
 DB 139 ADYEFIEQVQASYPHONVFCALDANSAAEFQKMKLQCCLPYVNLVPTSYSGKH 198
 QY 201 SRLADNLCSLDLKSISQWYVNLGQDFPLKSNFELVSEKLKANNLETVKPPNS 260
 DB 199 NINLAHNMCKK--LESIRGMGYLMLLQNHVDITKSYVELDRIFELGCVN-----DV 249
 QY 261 KLERFTYHHELRVPEYVKLPITNISKEPAPNIOIFGSAFVLSQAFVKYIFNFI 320
 DB 250 FMSREIPERRRKLKMDKSLKLFRRDSKNS--DELEISSQSVHSSRAAVKLYEVD 308
 QY 321 VQDFP-AMSKDYSPDEHFATLIRVPGI--PGEISRSA--QDVSDLSKRLVW--NYE 375
 DB 309 LSIFIDQNMOTDYGADDEGFIETPQMSDLMKPHFTNDCIHNDIAITYISRLAWPIYL 368
 QY 376 GFYFPCIGSHLRVYCIYGAELRWLTKDGHWFANKEDSKYDPL-----IKCLA 425
 DB 369 DRAAHCAIGTVRHDCIFGIDDFRAI-----SKF---PILNMLPAFDYSIIECTA 415
 QY 426 EKLEEQORDWITLSEKLFMDRNL 449
 DB 416 ELL--HNRTYLGVDQK--EKNL 435

RESULT 13
 T321566
 hypothetical protein F30A10.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T31566
 R:Barlow, K.
 Submitted to the EMBL Data Library, October 1996
 A:Reference number: 219442
 A:Accession: T31566
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1486 <W>LL>
 A:Cross-references: EMBL:Z81072; PIDN:CA803022.1; GSPDB:GN00019; CESP:F30A10.4
 A:Experimental source: clone F30A10

C:Genetics:
 A:Gene: CESP:F30A10.4
 A:Map position: 1
 A:Introns: 19/3; 59/3; 93/2; 248/3; 300/1; 348/3; 385/2; 429/1

Query Match 9.2%; Score 219.5; DB 2; Length 486;
 Best Local Similarity 22.9%; Pred. No. 1.5e-08;
 Matches 102; Conservative 80; Mismatches 202; Indels 61; Gaps 17;

QY 15 KYEFLFLMLSLKLNVR--RLEPOKDIY-----LVEXSLSPVRNRYTHV 64
 DB 36 RIFITALLISSTILTYSKSMIPKFIYRLSKERAPLKHITRSTKRA--DYVTE 93
 QY 65 DEVRVNSGTYEOPPLEIGKSLERRDDIILEDDDVVAMTSDCDIYQTLRGYAOKLV 124
 DB 94 SRSRY-LNCARLIDGDVDESIDYVNGR---MKLDEEKLFLSDMCDSDIQ--NRIFRMP 147
 QY 125 SKEESFPIAYSLVYHKAIMWERLIHAIFYQNIYCIHYORKAPDFKVMNNLACFS 184
 DB 148 PEKLRPLFAFRNITGYIELOEVFLSYHPDNIFCIAMDSEKLUKSMRTMADCFE 207
 QY 185 NFETASKLEAVEYAHISRLQADNLCSLDLKSISQWYVNLGQDFPLKSNFELVSEK 244
 DB 208 NTVYLDREYDMDRAGHKQDAHFCDLKOILDE--HMSHAITLQNFDLITKSPKOLDS 265
 QY 245 KINGANML-----ETVAPNSKLERTYHHELRVPE--YKLPITNISK 289
 DB 266 ILNTYSIMGEFDYGFYSRRTFEDWTGAKKL--FKNE--QSVPLEIILKRLKSLNE 320
 QY 290 EAPPHNIQIFGSAFVLSQAFVKYIFNNSIYQDFPANSK--TYSPEHFATLI--RVP 346
 DB 321 -----VLSKVFLSLFKVNLQNYIKRFDRLTEGVDEMMATLENYL 365
 QY 347 GIPGEISRSQDYSDLOSTRVYKNNYEG--FFYSCGSHR--SVCIYGAELRWLKD 404
 DB 366 GLDQWESNC--TVAKEDITFQTHHLEQSDGLYQDCKSKMLRHSICVIGVFEOLSKS 424
 QY 405 GHWFANKFDSKYDPLIKCLAEKLE 429
 DB 425 PMVIANKVTANPDFGTLICVREMIK 449

RESULT 14
 T33384
 hypothetical protein H4IC03.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T33384
 R:Attelle, P.; Kramer, J.
 Submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid H4IC03.
 A:Reference number: 221334
 A:Accession: T33384
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1472 <L>AT>
 A:Cross-references: EMBL:AF077545; PIDN:AA26306.1; GSPDB:GN00020; CESP:H4IC03.3
 A:Experimental source: strain Bristol N2; clone H4IC03
 C:Genetics:
 A:Gene: CESP:H4IC03.3
 A:Map position: 2
 A:Introns: 48/1; 94/3; 117/3; 220/1; 307/1; 354/3; 394/2; 406/1; 442/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein H4IC03.3

Query Match 8.5%; Score 203.5; DB 2; Length 472;
 Best Local Similarity 20.9%; Pred. No. 2.2e-07;
 Matches 90; Conservative 80; Mismatches 189; Indels 71; Gaps 16;

QY 20 FLTLWLISLKLNVRLF---PQKDIYLVEXSLSTSPFVRNRTYHKDEVREYVNSG 72
 DB 19 FKRIWILLFLAVCFVAYILMFPKGIY---DPLTTDQNMPTIBQYVNDLRNEIDQDN 75

```

OY 73 -----CSGIYEQP-----LEIGKSLIERRRDIIDLEDDVYAMTSDC 110
DB 76 AIISELQDFVMEKTEKTERKAIYPRPETDHIDCGRIILADKRVYLQTVSGADRYKIVENK 135
OY 111 DIYOTFLRGYACKLKEEKESFP-----IAYSLVHKDAIMVERLHAIVNHNHICIHDR 166
DB 136 NIMMCSAAMNRIIPSGNDKRLKNGVAFARIVTYDEMIKQVMSHPQNSFCFALDK 195
OY 167 KAPDTFFKAMNNLACFSNIFLASKLEAVEYAHISRLQADLCSLQWKYVYNL 226
DB 196 KAPSOFEHRLRAMAACLPVLLLPDEESYDSAGHNINLAHYNCLFLINKP-GWYVAILL 254
OY 227 CGQDPELKSNEFLVSELKLNKANMLEYKPPNSKLERFTYHHELRVRYEVYKPI--- 283
DB 255 QNHDIYVSVYIEQIYDMLGANDIE-ITPAGRVN-----KFKMDPVSLKMERN 306
OY 284 RTNISKAPPHNIQIFVGSAYFVLSQAFVKYLFNNSIVQDEF-AWSKDTYSPDEHFMTL 342
DB 307 ETGIDKNVLTSMKFAKGAOGLSRGAVDMVFTVDLSTYINQWNEGSSGVDEQFIQS- 365
OY 343 IRVP---GIPGEIS-----RSADVSDLOSKTRLVKMY---YEGFFYP--SQTGSHL 387
DB 366 FOVSADLGMFGHFTDECLKOKRNTDVS-----RMSQMYGSAHKSFDYSIVECTAELL 419
OY 388 RSVCIYGAEL 397
DB 420 YNRTFLGQVD 429

```

```

RESULT 15
T24745
hypothetical protein T09E11.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24745
R:McLay, K.
submitted to the EMBL Data Library, October 1996
A:Reference number: 219930
A:Accession: T24745
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1753 <util>
A:Cross-References: EMBL:Z81147; PIDN:CAB03539.1; GSPDB:GND0019; CESP:T09E11.6
A:Experimental source: clone T09E11
A:Gene: CESP:T09E11.6
A:Map position: 1
A:Introns: 34/2; 54/2; 191/3; 275/3; 315/3; 359/3; 380/3; 403/3; 426/3; 530/1; 617/1; 66

```

```

Query Match 8.5%; Score 202.5; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 4.9e-07;
Matches 96; Conservative 76; Mismatches 207; Indels 61; Gaps 16;

```

```

OY 21 LTFMLL-----SLKLINTRLFPKDIY--LYEYSLS-----TSPFVRNRYTHVKDEV 67
DB 335 LNRMLLITIAASIFSLVILLLEKEEPYQTLSESTLSCDEGPTKODLRRHTDIADYV 394
OY 68 R--YEVNCSGIYEQP---LEIGKSLIERRRDIIDLEDDVYAMTSDCIYQTLRGYA 120
DB 395 RERGFADKRGSYRRRPFANVDCGRILAGDKPYLOSLTGTRKRVKIVENCNLMNSCKAIR 454
OY 121 OKYSKEEK-----SFFIAYSLVHKDAIMVERLHAIVNHNHICIHDRKAPDEKVA 175
DB 455 SRLPSNDNITLRPLKHGIAFARIVTYKDYEFIEKQOVVSHPONAFCFVIDINASEFEKFR 514
OY 176 MNNLACFSNIFLASKLEAVEYAHISRLQADLCSLQWKYVYNLCGQDFPLKS 235
DB 515 MRALACMPRVIVLADDEDPYSSGHNVNLYHNKCLKALLDIP-GWYVALLQNHDLIMKS 573
OY 236 NFEVLVSELKLNKANMLEYKPPNSKLERFTYHHELRVRYEVYKLP-PI-----RTNI 287

```

```

DB 574 VYEMEQIFEWLGANDIEV-----HEIGRVYKRLKMDPMSLKLFINETEM 620
OY 288 SKAAPPHNIQIFVGSAYFVLSQAFVKYLFNNSIVQDEF-AWSKDTYSPDEHFMTLIRVP 346
DB 621 DKLLITPMKIVKGWVCSLSRASVEMMFQKLDPSIFMQNLNGRGYGVDEQYPIIQANA 680
OY 347 --GIPGEIS--RSADVSDLOSKTRLVKMYEGFFYP--SC-TGSHLRVCIYGAELRW 400
DB 681 EFGMRGHFTDECLQOGKTEFTETRIALM-----VPESKCDINMTRHANCIIGLEHFQA 733
OY 401 LKDGHWPFANKSDKVDPI 420
DB 734 VASFTHLMFNKKVSSSLDELI 753

```

```

Search completed: August 2, 2001, 14:45:50
Job time: 136 sec

```

Fri Aug 3 09:38:10 2001

us-09-645-192-2.rpt

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2001, 14:45:20 ; Search time 16.71 Seconds
(without alignments)
928.630 Million cell updates/sec

Title: US-09-645-192-2

Sequence: 1 MKIFKCYRKHTLQKVFILF.....DWITLPSEKLFMDRLTTTS 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	863.5	36.1	428	1 G6NT_HUMAN	Q02742 homo sapien
2	848.5	35.5	428	1 G6NT_MOUSE	Q09324 mus musculu
3	840	35.2	427	1 G6NT_BOVIN	Q92180 bos taurus
4	730.5	30.6	400	1 BGIB_MOUSE	P97402 mus musculu
5	711	29.8	400	1 BGIB_HUMAN	Q06430 homo sapien
6	711	4.9	895	1 YAS4_YEAST	Q38454 methanococ
7	108.5	4.5	661	1 WHI3_YEAST	Q34761 saccharomyc
8	102.5	4.3	654	1 BFR2_HUMAN	Q01742 homo sapien
9	102.5	4.3	821	1 FGR2_HUMAN	P21802 homo sapien
10	102.5	4.3	821	1 FGR2_MOUSE	P41653 mus musculu
11	101.5	4.2	2054	1 YCF2_PINTH	P09976 nicotiana t
12	101.5	4.2	2280	1 YCF2_TOBAC	Q08807 bacillus t
13	100.5	4.2	2607	1 BACB_BACLI	P21093 bacillus t
14	99.5	4.2	666	1 VO01_VACCC	P18461 gallus gall
15	99.5	4.2	823	1 CEK3_CHICK	Q24093 drosophila
16	99.5	4.1	1639	1 A23D_DROME	P04958 drosophilu
17	97.5	4.1	1314	1 TETX_CLOTE	P00591 homo sapien
18	97	4.1	440	1 GAP_HUMAN	O09255 ceenorhabd
19	97	4.1	1036	1 YOB6_CAEEL	Q35457 homo sapien
20	95.5	4.0	936	1 MSH4_HUMAN	Q92486 helioobacte
21	95	4.0	701	1 SYGB_HELPJ	P29434 buchnera ap
22	94	3.9	364	1 DP3B_MYCCE	P47247 mycoplasma
23	94	3.9	454	1 DNA3_BUCAP	P23116 mus musculu
24	94	3.9	1344	1 IF3A_MOUSE	O00610 homo sapien
25	94	3.9	1675	1 CIH1_HUMAN	P49951 bos taurus
26	94	3.9	1675	1 CIH_BOVIN	P14442 rattus norv
27	94	3.9	1675	1 CIH_RAT	O08006 b bacteraci
28	94	3.9	5255	1 BACA_BACLI	P51381 porphyra pu
29	93.5	3.9	884	1 SECA_PORPU	P38110 saccharomyc
30	93.5	3.9	2787	1 TEL1_YEAST	O9X528 ovis aries
31	93	3.9	503	1 CPV1_SHEEP	O60270 methanococ
32	93	3.9	520	1 Y208_METYA	P53304 drosophila
33	93	3.9	576	1 POLO_DROME	

34	92	3.9	908	1 DP01_BORBU	O51498 borrelia bu
35	92	3.9	1103	1 CYGF_BOVIN	O02740 bos taurus
36	92	3.9	1382	1 IF3A_HUMAN	Q14152 homo sapien
37	92	3.9	2022	1 ANTI_ONCVO	P21249 onchocerca
38	91.5	3.8	720	1 SPOT_MYCCE	P47520 mycoplasma
39	91.5	3.8	757	1 RTS1_YEAST	P38903 saccharomyc
40	91.5	3.8	1956	1 ATX1_PLAFA	O04956 plasmodium
41	91	3.8	344	1 Y613_METYA	O58030 methanococ
42	91	3.8	458	1 Y613_EUGGR	P05728 euglena gra
43	91	3.8	759	1 TEAL_YEAST	P47988 saccharomyc
44	91	3.8	809	1 LER1_BACAN	P15917 bacillus an
45	91	3.8	993	1 NISB_LACLA	P20103 lactococcus

ALIGNMENTS

RESULT 1	ID	Sequence	STANDARD:	PRT:	428 AA.
G6NT_HUMAN	002742;	01-JUN-1994 (Rel. 29, Last sequence update)			
G6NT_HUMAN	002742;	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	BETA-1,3-GALACTOSYL-O-GLYCOPROTEIN BETA-1,6-N-				
DE	ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)				
DE	(CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).				
GN	G6NT.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93028457; PubMed=1329093;				
RA	Bierhuizen M.F.A.; Fukuda M.;				
RT	"Expression cloning of a cDNA encoding UDP-GlcNAc:gal beta				
RT	1-3-galNAc-R (GlcNAc to GalNAc) beta-1-6-glcNAc transferase by gene				
RT	transfer into CHO cells expressing polyoma large tumor antigen."				
RL	Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISTUE=Placenta;				
RX	MEDLINE=96078409; PubMed=7579796;				
RA	Bierhuizen M.F.A.; Maemura K.; Kudo S.; Fukuda M.;				
RT	"Genomic organization of core 2 and I branching beta-1,6-N-				
RT	acetylglucosaminyltransferases. Implication for evolution of the				
RT	beta-1,6-N-acetylglucosaminyltransferase gene family."				
RL	Glycobiology 5:417-425(1995).				
CC	- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.				
CC	- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-				
CC	GALACTOSYL-1,3-(N-ACETYL-D-GLUCOSAMINYL-1,6)-N-ACETYL-D-				
CC	GALACTOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,6)-N-ACETYL-D-				
CC	GALACTOSAMINYL-R.				
CC	- PATHWAY: GLYCOSYLATION.				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.				
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES				
CC	AND MYELOID CELLS.				
CC	- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL: M97347; AAA35919.1; -				
CC	EMBL: L41415; AAA96661.1; -				
CC	PIR: A46293; A46293.				
CC	MIM: 600391; -				
CC	Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;				

KM Golgi stack; Glycoprotein.
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 10 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT
 FT DOMAIN 33 428 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 428 AA; 49784 MW; 7C9CEB51DA0A60 CRC64;

Query Match 36.1%; Score 863.5; DB 1; Length 428;
 Best Local Similarity 43.1%; Pred. No. 2e-55;
 Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

QY 28 LKLLNVRRLF--PQKDIYLV-EYSLSSTPFRNRYTHVDE--VRY-----EV 71
 DB 1 MLRLRLRLFSYPTKRYFNVLSLTFESVLR---IHQKPEFVSRLHLEAGNPSDI 57
 QY 72 NCSGIYDEPLEIGK-----SLEIRRDIILEDVVAMTSDDIYQTLRGYAOKLYS 125
 DB 58 NCTVYLOGDVNEIOKVKLEILTVKFKRP--RMTPDYIMTSDCSSFKRRKYIPELS 115
 QY 126 KEKSPFIATSLVYHKDAIWERLTHAIYNOHNYCIHYRKAAPDFKYAMNLLACFSN 185
 DB 116 KEKSPFIATSIYVHKHIEMLDRLRAIYMPQFYCHVDTKSEDEYLAAMGIAFCFSN 175
 QY 186 IFIAKLEAVEYAHISRLQADLNCSDLKSSIQMKYVNLGQDFPLKSNFEVSELK 245
 DB 176 VFVASRLESVYVYASWSRVQADLNCMDLAMSANWKYLMCGMDPEPKINLEIVRKLK 235
 QY 246 LMGANMLETVKPPSKLEFTHYHHLRVRPEYV--KLPIRTNISKEAPPHNIQIFVSGA 303
 DB 236 LMGNNLETERMSHKEERW-----KKREYVNGKL-TNTGTVMPLPLEPLESGSA 287
 QY 304 YFVLSQAFVKYIFNNSIYODFFANSKDTPSPDEHFMTLLRVPGIPGELISRSQA-DVSD 362
 DB 288 YFVVSREYGVYVQNEKIQKLMWAQDTPSPDEYLMATIORIPEVPSGLPSHSHYDLSDM 347
 QY 363 OSKRILYKMYVEGEF-----YPSCTGSHLSVCIYGAELRWLIKDGHPFANKFDSKVD 417
 DB 348 QAVAFPKWQYFEGDVSKGADYPPCDGVHVSVCIFSGADLMMLRKHHLPANKFEDVD 407
 QY 418 PILIKCIAEKLEEQ 431
 DB 408 LFAIQCLDEHLRHK 421
 RESULT 2
 G6NT_MOUSE STANDARD; PRT; 428 AA.
 ID G6NT_MOUSE
 AC Q09324;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-
 DE ACETYLGALUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)
 DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT).
 GN G6NT1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN NCBI_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN-DBA/2;
 RA Warren C.E., Smookler D.S., Dennis J.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCAN.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
 CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-
 CC GALACTOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,6)-N-ACETYL-D-
 CC GALACTOSAMINYL-R.

CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).

CC EMBL: U19265; AAA60948.1; -

DR MGD; MGI:95676; G6NT1.

KM Transferase; Glycosyltransferase; Transmembrane; Signal anchor;

KM Golgi stack; Glycoprotein.
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 10 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT
 FT DOMAIN 33 428 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 58 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 428 AA; 49649 MW; F3FED2517B27A81 CRC64;

Query Match 35.5%; Score 848.5; DB 1; Length 428;
 Best Local Similarity 42.4%; Pred. No. 2.4e-54;
 Matches 183; Conservative 72; Mismatches 138; Indels 39; Gaps 11;

QY 28 LKLLNVRRLF--PQKDIY-LVEYLSSTPFRNRYTHVDE--VRY-----EV 71
 DB 1 MLRLNFRRLFCPTKRYFMLVLSLTFESVLR---IHQKPEFVSRLHLEAGDPPYSNV 57
 QY 72 NCSGIYDEPLEIGK-----SLEIRRDIILEDVVAMTSDDIYQTLRGYAOKLYS 125
 DB 58 NCTKLIQGPBEIOKVKLEILTVQFKRP--RMTPHDYIMTRDCASTRIRKRYIPELT 115
 QY 126 KEKSPFIATSLVYHKDAIWERLTHAIYNOHNYCIHYRKAAPDFKYAMNLLACFSN 185
 DB 116 KEKSPFIATSIYVHKHIEMLDRLRAIYMPQFYCHVDTKSEDEYLAAMGIAFCFSN 175
 QY 186 IFIAKLEAVEYAHISRLQADLNCSDLKSSIQMKYVNLGQDFPLKSNFEVSELK 245
 DB 176 VFVASRLESVYVYASWSRVQADLNCMDLAMSANWKYLMCGMDPEPKINLEIVRKLK 235
 QY 246 LMGANMLETVKPPSKLEFTHYHHLRVRPEYV--KLPIRTNISKEAPPHNIQIFVSGAV 305
 DB 236 STGNNLETERMSHKEERW-----KRYAVVDGKL-TNTGTVMPLPLEPLESGSAVF 289
 QY 306 YFVLSQAFVKYIFNNSIYODFFANSKDTPSPDEHFMTLLRVPGIPGELISRSQADVSDLS 364
 DB 290 YFVREYGVYVQNEKIQKLMWAQDTPSPDEYLMATIORIPEVPSGLPSHSHYDLSDMA 349
 QY 365 KTRLVKMYVEGEF-----YPSCTGSHLSVCIYGAELRWLIKDGHPFANKFDSKVDPI 419
 DB 350 IAFVAKWQYFEGDVSAGYPPCDGVHVSVCIFSGADLMMLRKHHLPANKFEDVDVDF 409
 QY 420 PILIKCIAEKLEEQ 431
 DB 410 AIOCLDEHLRHK 421
 RESULT 3
 G6NT_BOVIN STANDARD; PRT; 427 AA.
 ID G6NT_BOVIN
 AC Q92180;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-
 DE ACETYLGALUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)
 DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).
 GN G6NT1.

OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Li C.M., Cheng P.W.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
 CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-
 CC GALACTOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,6)-N-ACETYL-D-
 CC GALACTOSAMINYL-R.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U41320; AAB3244.1; -
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack; Glycoprotein.
 RN DOMAIN 1 9
 FT TRANSMEM 10 32
 FT (POTENTIAL).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 1 9
 FT (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 52 427
 FT 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 58
 FT 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 95 95
 FT 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 427 AA; 49733 MW; 41B9CBF94BD4196 CRC64;
 Query Match 35.2%; Score 840; DB 1; Length 427;
 Best Local Similarity 42.8%; Pred. No. 9,8e-54;
 Matches 185; Conservative 73; Mismatches 134; Indels 40; Gaps 12;

DB 409 AIOCLDEHMKR 420
 RESULT 4
 BG1E MOUSE
 ID BG1E MOUSE STANDARD: PRT; 400 AA.
 AC P97402;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE N-ACETYLGLUCOSAMINIDE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE
 DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)
 DE (IGNT) (LARGE I ANTIGEN-FORMING BETA-1,6-N-
 DE ACETYLGLUCOSAMINYLTRANSFERASE).
 GN GCNT2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=9728061; PubMed=9134435;
 RX Magnet A.D.; Fukuda M.;
 RT "Expression of the large I antigen forming beta-1,6-N-
 RT acetylglucosaminyltransferase in various tissues of adult mice";
 RL Glycobiology 7:285-295(1997).
 CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED
 CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
 CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED
 CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL
 CC -1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-
 CC GLUCOSAMINYL-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U68182; AAB3962.1; -
 DR MGD; MGI:1100870; Gcnt2.
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack; Glycoprotein.
 RN DOMAIN 1 6
 FT TRANSMEM 7 25
 FT (POTENTIAL).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 1 6
 FT (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 26 400
 FT 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 37 37
 FT 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255
 FT 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314
 FT 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388
 FT 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 400 AA; 45507 MW; 193FF3862E3B9691 CRC64;
 Query Match 30.6%; Score 730.5; DB 1; Length 400;
 Best Local Similarity 40.0%; Pred. No. 7,6e-46;
 Matches 156; Conservative 51; Mismatches 160; Indels 23; Gaps 6;

```

QY 163 HYDRAPDTEFKVAMNNLAKCFNSIFIAASKLEAVEYAHISRLQADLNCISDLKSSIQWKY 222
DB 125 HVEKRTAEKGAEDVLSQFEPVFLASKMEPVYGGISRLQADLNCISDLSTSEVPWKY 184
QY 223 VINLQSGDPEPLKSNFELVSELKINGANMLETVKPPNSKLERFTY--HHBLRRVPEYVKL 281
DB 185 AINTCODEPFLKTNKEIVQYKGLKGNKILPGLVLPRAHAIGRTRYVHREHLSKELSYV-- 242
QY 282 PITNISKAPPHNIQIFVGSATFVLSQAFVKIFNNISIVODEFANSKDTYSPDEHFMAT 341
DB 243 -IRFTALKPPPHNLTITGSAVALSREFANFVLDPRAVDLHWSKDTSPDEHFMAT 301
QY 342 LIRVPGIPGIEISSADVDSDLSQKTRLVKNNYEGEFPYPSCTGSHLSRVCIGAAELRWL 401
DB 302 LNIIPGVGSPPMAS-----WTGNLRAVKMDMEK-AKHGCCGHYVHGICITGNGDLQWL 355
QY 402 IKDGHFANKFDSKVDPIKICLAEKLEQ 431
DB 356 INSOSLEFANKFELNTYPLTVECLELRER 385

RESULT 5
BG1B_HUMAN STANDARD; PRT; 400 AA.
AC 006450;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE N-ACETYLGLUCOSAMINIDE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE
DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (1-BRANCHING ENZYME)
DE (ICGTN).
GN GCNT2 OR IT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93194065; PubMed=8449405;
RA Bierhuizen M.F.A., Mattel M.-G., Fukuda M.;
RT "Expression of the developmental I antigen by a cloned human cDNA
RT encoding a member of a beta-1,6-N-acetylglucosaminyltransferase gene
RT family."
RL Genes Dev. 7:468-478(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96078409; PubMed=7579796;
RA Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT "Genomic organization of core 2 and I branching beta-1,6-N-
RT acetylglucosaminyltransferases. Implication for evolution of the
RT beta-1,6-N-acetylglucosaminyltransferase gene family."
RL Glycobiology 5:417-425(1995).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=98070745; PubMed=9405606;
RA Sasaki K., Kurata-Miura K., Ujita M., Angata K., Nakagawa S.,
RA Sekine S., Nishi T., Fukuda M.;
RT "Expression cloning of cDNA encoding a human beta-1,3-N-
RT acetylglucosaminyltransferase that is essential for poly-N-
RT acetylglucosamine synthesis."
RL Proc. Natl. Acad. Sci. U.S.A. 94:14294-14299(1997).
CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED
CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED
CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
CC -1,4-N-ACETYL-D-GLUCOSAMINYLR- = UDP + N-ACETYL-BETA-D-
CC GLUCOSAMINYLR-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYLR.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, HIGHLY EXPRESSED IN PROSTATE AND
CC TO A LESSER EXTENT IN SMALL INTESTINE AND COLON. BARELY DETECTED

```

```

CC CC IN HEART, BRAIN, KIDNEY AND PANCREAS. NO EXPRESSION IN PLACENTA,
CC CC LUNG, LIVER, SKELETAL MUSCLE, SPLEEN, THYMUS, TESTIS, OVARY AND
CC CC PERIPHERAL BLOOD LEUKOCYTES. IN FETUS, HIGHLY EXPRESSED IN BRAIN
CC CC AND TO A LESSER EXTENT IN LUNG AND KIDNEY. BARELY DETECTED IN
CC CC LIVER.
CC CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION INCREASES DRAMATICALLY DURING
CC CC DEVELOPMENT AND ONCOGENESIS.
CC CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb.ch/announce/
CC CC or send an email to license@isb.ch).
CC CC -----
DR EMBL; L19659; AAA81777.1; -
DR EMBL; Z19550; CA97610.1; -
DR EMBL; L41607; AAA99832.1; -
DR EMBL; L41605; AAA99832.1; JOINED.
DR EMBL; L41606; AAA99832.1; JOINED.
DR PIR; A46297; A46297.
DR MIM; 600429; -
DR MIM; 110800; -
KM Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
KM Transferrase; Glycoprotein.
FT FT DOYAIN 1 6
FT FT TRANSMEM 7 25
FT FT CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT FT LUMENAL, CATALYTIC (POTENTIAL).
FT FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 26 400
FT FT CARBOHYD 37 37
FT FT CARBOHYD 255 255
FT FT CARBOHYD 314 314
FT FT CARBOHYD 388 388
SQ SEQUENCE 400 AA; 45854 MW; 1469628690A1F43B CRC64;

Query Match 29.8%; Score 711; DB 1; Length 400;
Best Local Similarity 43.7%; Pred. No. 2e+44;
Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SCDIYQTLRGYAOGLVSKESKSPFIAYSLVYHNDKAIWERTYHAIYQNHNYICLHDK 167
DB 70 SSKCEYLQSHYITAPLSSEADPFLAYIMVTHHFDFAFLRAIYPMQNYCVHDEK 129
QY 168 ADPTKRVAMNNLAKCFNSIFIAASKLEAVEYAHISRLQADLNCISDLKSSIQWKY 227
DB 130 ATTERDAVEQLSCFPNPAFLASKMEPVYGGISRLQADLNCISDLSTSEVPWKY 189
QY 228 GODFPLKSNFELVSELKINGANMLETVKPPNSKLERFTY--HHBLRRVPEYVKLPIRTN 286
DB 190 GODFPLKTNKEIVQYKGLKGNKILPGLVLPRAHAIGRTRYVHREHLSKELSYV--IRTT 246
QY 287 ISKEAPPHNIQIFVGSATFVLSQAFVKIFNNISIVODEFANSKDTYSPDEHFMATLIRVP 346
DB 247 AKRPPPHNLTITGSAVALSREFANFVLDPRAVDLHWSKDTSPDEHFMATLIRIP 306
QY 347 GIPGIEISSADVDSDLSQKTRLVKNNYEGEFPYPSCTGSHLSRVCIGAAELRWLINDGH 406
DB 307 GYFGSPPMAS-----WTGNLRAIKWMDMEK-RHGCGHGHYVHGICITGNGDLKMLVNSPS 360
QY 407 WFANKFDSKVDPIKICLAEKLEQ 433
DB 361 LFANKFELNTYPLTVECL--ELRRER 385

RESULT 6
YAS4_METJA STANDARD; PRT; 895 AA.
AC 058454;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)

```


[2]
RN SEQUENCE FROM N.A.
RP STRAIN-S288C / F.Y1679;
RC MEDLINE=95242839; PubMed=7725799;
RX
RA Jomiaux J.-L., Coster F., Purnelle B., Goffeau A.;
RT "A 21.7 kb DNA segment on the left arm of yeast chromosome XIV
RT carries WRI3, GCZ2, SPX18, SPX19, an homologue to the heat shock gene
RT SSBI and 8 new open reading frames of unknown function.";
RL Yeast 10:1639-1645(1994).
CC - FUNCTION: INVOLVED IN SIZE CONTROL; POSSIBLE RNA BINDING PROTEIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way commercially modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-slb.ch/announce/>)

```

DR EMBL: U01095: AAA03320.1; -
DR EMBL: X78898; CAA55511.1; -
DR EMBL: Z71473; CAA96092.1; -
DR HSSP: P09012; 201A.
DR SGD: S0005141; WHI3.
DR InterPro: IPR000504; -
DR Pfam: PF00076; rrm; 1.
DR PROSITE: PS50102; rrm; 1.
DR KMW RNA-binding.
FT DOMAIN 247
SQ SEQUENCE 661 AA; 71253 MW; 476AB9716CF96043 CRC64;

Query Match 4.5%; Score 108.5; DB 1; Length 661;
Best Local Similarity 20.0%; Pred. No. 1.6;
Matches 68; Conservative 60; Mismatches 125; Indels 87; Gaps 14

QY 117 KGYAKLVKKEKSFPIASVLYHKAI--MYERLHAIYNQHNICIHVDRKAPQPFKY 174
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 52 RSAGSELKYLNDTNSLALISMLNTLALSGMPDIASSNISN-----DNNTKGSYSL 104
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 AMNNLAK-----CFENIFASKLEAVEY---AHISRLQADINCLSD-----LTKS 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 105 KLSNVAKDTLKECYIIFALAEVSKSIELQKKSSSSITSASLEDDENDIFIAFELNL 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 217 SIQWKYVIN---LCQDPELKSNEFLVSELK-----LNGANNLETYK 256
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 165 AIIVYAVILSKNNELFEPSPNKTVEIIDDTKNTLVSPSSAIFNDTSRLKNSN--SGMK 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 PPN--SKLERFY-----HHELRRPVYIVYALPIRTINSKAPPNIDQIFVGSAY 304

```

```

Db 223 RPSLQSRARFSPSPFNSDPLSQQQSQQQPQPOGHSTOKHSPQCCNQGVNSI 282
QY 305 FVLQAFVKYIFNNSIVQDFPWS-----KUTYSPDEHFATLIRVPGIGE 351
Db 283 PLSSQGVIGLHNSHSHQDLVSVESTIOTSDIGKSPLLRNTEINERKWT-----GSPSS 338
QY 352 ISRAQDVSDLSQKRLVKW-----NYEGFFPYSCTS 385
Db 339 INGV---MSTPQSTPLEMGNTSASQHGSSFLPSAST 375

RESULT 8
BFR2_HUMAN
ID BFR2_HUMAN STANDARD; PRT; 654 AA.
AC 001742;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FIBROBLAST GROWTH FACTOR RECEPTOR BFR-2 PRECURSOR (EC 2.7.1.112).
GN BFR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=91274356; PubMed=1647213;
RA Seno M., Sasada R., Watanabe T., Ishimaru K., Igarashi K.;
RT "Two CDNA's encoding novel human FGF receptor.";
RL Biochim. Biophys. Acta 1089:244-246(1991)
CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH
CC FACTORS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X56191; NOT_ANNOTATED_CDS.
DR HSSP: P11362; IFG1.
DR InterPro: IPR000719;
DR InterPro: IPR001245;
DR InterPro: IPR003006;
DR Pfam: PF00047; Ig_2;
DR Pfam: PF00069; Pkinase_1.
DR PRINTS: PR00109; TYRKINASE.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW Transferrase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW Signal.
FT SIGNAL 1 21
FT CHAIN 22 654
FT DOMAIN 22 263
FT TRANSMEM 264 284
FT DOMAIN 285 654
FT DOMAIN 57 123
FT DOMAIN 156 234
FT DOMAIN 367 647
FT NP_BIND 373 381
FT BINDING 403 403
FT ACT_SITE 512 512
FT MOD_RES 543 543
FT CARBOHYD 113 113

```

```

FT CARBOHYD 126 126
FT CARBOHYD 150 150
FT CARBOHYD 182 182
FT CARBOHYD 203 203
FT CARBOHYD 214 214
FT CARBOHYD 235 235
SQ SEQUENCE 654 AA; 73594 MW; F4954E7DC70FD233 CRC64;

Query Match 4.38; Score 102.5; DB 1; Length 654;
Best Local Similarity 24.08; Pred. No. 4.2; Mismatches 103; Gaps 18;
Matches 78; Conservative 35;

QY 182 CFSNFIASKL-----EAVEXAHISRLQADL--NCLSDL-----LKSSTQMKYVIN 225
Db 377 CFCGVVAEAVGIDKDKREAVTVA-VKMLKDDATEKDLSDVSEMMKMKIKHKNIN 435
QY 226 L---CGGFPIKSNFELVSELKINGANMLETVPKPSKLEFTHHELRVPEYVKLP 282
Db 436 LLGACTQDGPL--YVIVEYASKGNLREYLRARPQGE-----YSTDKNVPEQMTF- 486
QY 283 IRTNISKAPPNIIQIEYGSAYFY-----LSQAFVKYIFNNSIVQDFPWSKDTYSP 334
Db 487 -----KLVSGTYQLARGMEYIASQKCIHRDLARVLYENVMKIADEGLARDINNI 540
QY 335 DEHFATLIRVPGIRGEISRAQDVSDLSQKRLVKRWYVYGEFFPYSCTSGLRSYCIY 394
Db 541 DYKKTTRGRLP-----VKMAPALDLYVT--HQSDVWSFG 576
QY 395 -----AELRLKDGHWFKANKFDSKVDPIILIKLAEKLEQQRD-W 435
Db 577 VLMETFTLGSPYIGIVVELEFKLKSGH---RMDKPN-----CTNE-LYMMMDCW 626
QY 436 IILPSEK-LF-----MDRNLTTTS 453
Db 627 HAVPSQRPFTKQVLEDRLITLT 651

RESULT 9
BFR2_HUMAN
ID BFR2_HUMAN STANDARD; PRT; 821 AA.
AC P21802; P18443;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FIBROBLAST GROWTH FACTOR RECEPTOR 2 PRECURSOR (FGFR-2) (EC 2.7.1.112)
DE (KERATINOCYTE GROWTH FACTOR RECEPTOR).
GN FGFR2 OR BEK OR BFR-1 OR KSAM-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=91045561; PubMed=2172978;
RA Houssaint E., Blanquet P.R., Champion-Arnaud P., Gesnel M.C.,
RA Torriglia A., Courtois Y., Breathnach R.;
RT "Related fibroblast growth factor receptor genes exist in the human
RT genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8180-8184(1990).
RP SEQUENCE FROM N.A.
RP MEDLINE=91045561; PubMed=2172978;
RP Houssaint E., Blanquet P.R., Champion-Arnaud P., Gesnel M.C.,
RP Torriglia A., Courtois Y., Breathnach R.;
RP "Related fibroblast growth factor receptor genes exist in the human
RP genome.";
RP Proc. Natl. Acad. Sci. U.S.A. 87:8180-8184(1990).
RX MEDLINE=91274356; PubMed=1647213;
RX Seno M., Sasada R., Watanabe T., Ishimaru K., Igarashi K.;

```

RT "Two CDNA's encoding novel human EGF receptor.";
 RL Biochim. Biophys. Acta 1089:244-246(1991).
 RN [14]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Stomach cancer;
 RX MEDLINE=90332706; PubMed=2377625.
 RA Hattori Y., Odagiri H., Nakatani H., Miyagawa K., Naito K.,
 RA Sakamoto H., Katoh O., Yoshida T., Sugimura T., Terada M.,
 RT "K-sam, an amplified gene in stomach cancer, is a member of the
 RT heparin-binding growth factor receptor genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5983-5987(1990).
 RN [15]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92212948; PubMed=1313574;
 RA Katoh M., Hattori Y., Sasaki H., Tanaka M., Sugano K., Yazaki Y.,
 RA Sugimura T., Terada M.;
 RT "K-sam gene encodes secreted as well as transmembrane receptor
 RT tyrosine kinase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:2960-2964(1992).
 RN [16]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=93016048; PubMed=1400433;
 RA Dell K.R., Williams L.T.;
 RT "A novel form of fibroblast growth factor receptor 2. Alternative
 RT splicing of the third immunoglobulin-like domain confers ligand
 RT binding specificity.";
 RL J. Biol. Chem. 267:21225-21229(1992).
 RN [17]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Blood;
 RA Steinberger D., Mueller U.;
 RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [18]
 RP VARIANTS CS HIS-340; ARG-342; SER-342; TYR-342 AND CYS-354.
 RX MEDLINE=95078932; PubMed=7987400;
 RA Reardon W., Winter R.M., Rutland P., Pulley L.J., Jones B.M.,
 RA Malcolm S.;
 RT "Mutations in the fibroblast growth factor receptor 2 gene cause
 RT Crouzon syndrome.";
 RL Nat. Genet. 8:98-103(1994).
 RN [19]
 RP VARIANTS CS CYS-328 AND CYS-347, AND VARIANT JWS GLY-344.
 RX MEDLINE=95179174; PubMed=7874170;
 RA Jabs E.W., Li X., Scott A.F., Meyers G., Chen W., Eccles M., Mao J.,
 RA Charas L.R., Jackson C.E., Jaye M.;
 RT "Jackson-Weiss and Crouzon syndromes are allelic with mutations in
 RT fibroblast growth factor receptor 2.";
 RL Nat. Genet. 8:275-279(1994).
 RN [10]
 RP VARIANTS CS.
 RX MEDLINE=95384152; PubMed=7655462;
 RA Rutland P., Wilkie A.O.M., Slaney S.F., Poole M.D., Pulley L.J.,
 RA Rutland P., Hockley A.D., Wake M.J.C., Goldin J.H., Winter R.M.,
 RA Reardon W., Malcolm S.;
 RT "Mutations in the third immunoglobulin domain of the fibroblast growth
 RT factor receptor 2 gene in Crouzon syndrome.";
 RL Hum. Mol. Genet. 4:1077-1082(1995).
 RN [11]
 RP VARIANT PS ALA-321.
 RX MEDLINE=95235551; PubMed=7719333;
 RA Lajeune E., Wei M.H., Bonaventure J., Munnich A., Le Merrer M.,
 RA Renier D.;
 RT "FGFR2 mutations in Pfeiffer syndrome.";
 RL Nat. Genet. 9:108-108(1995).
 RN [12]
 RP VARIANTS AS TRP-252 AND ARG-253.
 RX MEDLINE=95235562; PubMed=7719344;
 RA Wilkie A.O.M., Slaney S.F., Oldridge M., Poole M.D., Ashworth G.J.,
 RA Hockley A.D., Hayward R.F., David D.J., Pulley L.J., Rutland P.,
 RA Malcolm S., Winter R.M., Reardon W.,
 RT "Apert syndrome results from localized mutations of FGFR2 and is
 RT allelic with Crouzon syndrome.";
 RL Nat. Genet. 9:165-172(1995).
 RN [13]
 RP VARIANTS PS PRO-341; ARG-342 AND TYR-342.
 RX MEDLINE=95235563; PubMed=7719345;
 RA Rutland P., Pulley L.J., Reardon W., Baralster M., Hayward R.,
 RA Jones B., Malcolm S., Winter R.M., Oldridge M., Slaney S.F.,
 RA Poole M.D., Wilkie A.O.M.;
 RT "Identical mutations in the FGFR2 gene cause both Pfeiffer and
 RT Crouzon syndrome phenotypes.";
 RL Nat. Genet. 9:173-176(1995).
 RN [14]
 RP VARIANTS CS/JWS/PS.
 RX MEDLINE=96203627; PubMed=8644708;
 RA Meyers G.A., Day D., Goldberg R., Daentl D.L., Prylepa K.A.,
 RA Abrams L.J., Graham J.M. Jr., Feingold M., Moeschler J.B.,
 RA Ramsley E., Scott A.F., Jabs E.W.;
 RT "FGFR2 exon IIIa and IIIC mutations in Crouzon, Jackson-Weiss, and
 RT Pfeiffer syndromes: evidence for missense changes, insertions, and a
 RT deletion due to alternative RNA splicing.";
 RL Am. J. Hum. Genet. 58:491-498(1996).
 RN [15]
 RP VARIANTS CS CYS-105; GLU-338; CYS-351 AND ARG-384.
 RX MEDLINE=97101656; PubMed=8946174;
 RA Pulley L.J., Reardon W., Wilkes D., Rutland P., Jones B.M.,
 RA Hayward R., Hall C.M., Brueton L., Chun N., Lammer E., Malcolm S.,
 RA Winter R.M.;
 RT "Spectrum of craniosynostosis phenotypes associated with novel
 RT mutations at the fibroblast growth factor receptor 2 locus.";
 RL Eur. J. Hum. Genet. 4:283-291(1996).
 RN [16]
 RP VARIANTS CS ILE-331; ASP-336--ALA-337 DUPL AND TRP-356--THR-358 DEL.
 RX MEDLINE=97114301; PubMed=8956050;
 RA Steinberger D., Mulliken J.B., Mueller U.;
 RT "Crouzon syndrome: previously unrecognized deletion, duplication, and
 RT point mutation within FGFR2 gene.";
 RL Hum. Mutat. 8:386-390(1996).
 RN [17]
 RP VARIANT PS CYS-290.
 RX MEDLINE=97295073; PubMed=9150725;
 RA Tartaglia M., Valeri S., Valardi F., di Rocco C., Battaglia P.A.;
 RT "Trp290Cys mutation in exon IIIa of the fibroblast growth factor
 RT receptor 2 (FGFR2) gene is associated with Pfeiffer syndrome.";
 RL Hum. Genet. 99:602-606(1997).
 RN [18]
 RP VARIANT CS L-252, VARIANT AS F-252, AND VARIANT PS F-252--S-253.
 RX MEDLINE=97156222; PubMed=9002682;
 RA Oldridge M., Lunt P.W., Zackai E.H., McDonald-Mcglin D.M., Nuenke M.,
 RA Moloney D.M., Twig S.R.F., Heath J.K., Howard T.D., Hoganson G.,
 RA Gagnon D.M., Jabs E.W., Wilkie A.O.M.;
 RT "Genotype-phenotype correlation for nucleotide substitutions in the
 RT Igit-Igit linker of FGFR2.";
 RL Hum. Mol. Genet. 6:137-143(1997).
 RN [19]
 RP VARIANT CS GLU-292.
 RX MEDLINE=97297373; PubMed=9152842;
 RA Steinberger D., Collmann H., Schmalenberger B., Mueller U.;
 RT "A novel mutation (a886g) in exon 5 of FGFR2 in members of a family
 RT with Crouzon phenotype and plagiocephaly.";
 RL J. Med. Genet. 34:420-422(1997).
 RN [20]
 RP VARIANTS CS VAL-276; CYS-301 AND SER-314.
 RX MEDLINE=98180879; PubMed=9521581;
 RA Steinberger D., Vriend G., Mulliken J.B., Mueller U.;
 RT "The mutations in FGFR2-associated craniosynostoses are clustered in
 RT five structural elements of immunoglobulin-like domain III of the
 RT receptor.";
 RL Hum. Genet. 102:145-150(1998).
 RN [21]
 RP VARIANTS AS TRP-252 AND ARG-253.
 RX MEDLINE=98112406; PubMed=9452027;
 RA Tsai F.-J., Hwu W.-L., Lin S.-P., Chang J.-G., Wang T.-R., Tsai C.-H.;
 RT "Two common mutations 934C to G and 937C to G of fibroblast growth
 RT factor receptor 2 (FGFR2) gene in Chinese patients with Apert

RT syndrome." Hum. Mutat. Suppl. 1:S18-S19(1998).

RL (22)

RN VARIANT PS CYS-351.

RX MEDLINE=98358420; PubMed=9693549;

RA Hesthissen J.M., Vaandrager J.M., Hoogeboom A.J.,

RT fibroblast growth factor receptor-2 gene."

RL J. Craniofac. Surg. 9:207-209(1998).

CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH

CC FACTORS.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- DISEASE: DEFECTS IN FGFR2 ARE A CAUSE OF CROUZON SYNDROME (CS).

CC ALSO CALLED CRANIOFACIAL DYSOSTOSIS TYPE I (CPD1). CHARACTERIZED

CC BY CRANIOSINOSTOSIS (PREMATURE FUSION OF THE SKULL SUTURES),

CC HYPERTELORISM, EXOPHTHALMOS AND EXTERNAL STRABISMOS, PARKOT-BEAKED

CC NOSE, SHORT UPPER LIP, HYPOPLASTIC MAXILLA, AND A RELATIVE

CC MANDIBULAR PROGNATHISM.

CC -1- DISEASE: DEFECTS IN FGFR2 ARE A CAUSE OF JACKSON-WEISS SYNDROME

CC (JWS) CHARACTERIZED BY CRANIOFACIAL ABNORMALITIES AND ABNORMALITY

Query Match 4.3%; Score 102.5; DB 1; Length 821;

Best Local Similarity 24.0%; Pred. No. 5.7; Mismatches 109; Gaps 18;

Matches 78; Conservative 35; Indels 103;

CC 182 CFSNFIASKL-----EAVEYAHISRLQADL--NCLSDL-----LKSSIQMKYVIN 225

DB 491 CFGGVMAEAVGIDKDKPKRAVYA-VKMLKDATEKDSLVSEMMMKHKHNIIN 549

QY 226 L---CGGDFPKSNFELVSELRKINGANMLEYKPPNSKLEERTYHHLRVRVEYVKLP 282

DB 550 LLAGCTODGDL---YVIVEASKGNLREYLARPPGME-----YSYDINRVEEDMTF- 600

QY 283 IRTNISEAPPNHIQIFVGSAYFY-----LSQAFVKYIFNNSIVQDFPANSKPTYSP 334

DB 601 -----KDIVSCTQIARCHEYILASOKCIHRDLARVYLTVENNVMKIAFGIARDINNI 654

QY 335 DEHFATLIRVYGIPEISRSADVSDLSQKTRLVKNVYEGFPYSCGSHLRSCYIG 394

DB 655 DYKKTNGRLP-----VKWMAPEALFDRVYT--HOSDWWSFG 690

QY 395 -----AELRWLIKQHWFKPFSKYDPILKLAELSEQORD-N 435

DB 691 VLMETITLGGSPYGPVPELEFKLEKGH---RMDKPN-----CTNE-LVMWMDCM 740

QY 436 ILLPSEK-LF-----MDRNLTTTS 453

DB 741 HAVDSORPTEKQVEDLDRLITLT 765

RESULT 10

FGFR2_MOUSE STANDARD; PRT; 821 AA.

ID P21803; Q61342; Rel. 18, Created)

AC 01-MAY-1991 (Rel. 35, last sequence update)

DT 01-NOV-1997 (Rel. 38, last annotation update)

DE FIBROBLAST GROWTH FACTOR RECEPTOR 2 PRECURSOR (FGFR-2) (EC 2.7.1.112)

EN (KERATINOCYTE GROWTH FACTOR RECEPTOR).

GN FGFR2 OR ECT1 OR BEK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A. (LONG FORM).

RC TISSUE=Brain, and Liver;

RX MEDLINE=9228773; PubMed=1373495;

RA Mansukhani A., Dell'Era P., Moscatelli D., Kornbluth S.,

RA Hanafusa H., Basilio C.;

RT "Characterization of the murine BEK fibroblast growth factor (FGF)

RT receptor: activation by three members of the FGF family and

RT requirement for heparin".

RL Proc. Natl. Acad. Sci. U.S.A. 89:3305-3309(1992).

RN (2)

RN SEQUENCE FROM N.A. (SHORT FORM).

RX MEDLINE=91095977; PubMed=1846048;

RA Miki T., Fleming T.P., Bottaro D.P., Rubin J.S., Ron D.,

RA Aaronson S.A.;

RT "Expression cDNA cloning of the KGF receptor by creation of a

RT transforming autocoding loop."

RL Science 251:72-75(1991).

RN [3]

RN SEQUENCE OF 477-821 FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=89219016; PubMed=2468999;

RA Kornbluth S., Paulson K.E., Hanafusa H.;

RT "Novel tyrosine kinase identified by phosphotyrosine antibody

RT screening of cDNA libraries."

RL Mol. Cell. Biol. 8:5541-5544(1988).

CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH FACTORS.

CC POSSESSES A HIGHER AFFINITY FOR ACIDIC THAN FOR BASIC FGF'S.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN (A LONG AND A

CC SHORT FORM) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME

CC GENE.

CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR

CC FAMILY.

CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE DOMAINS.

CC -----

CC This SWISS-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M86441; AAA37286.1; -

CC EMBL: M63503; AAA39377.1; -

CC EMBL: M23362; AAA37285.1; -

CC PIR: A31378; TVMSBK.

CC PIR: A38429; A38429.

CC HSP: P11362; IFG1.

CC MGI: MGI:95523; Fgf2.

CC InterPro: IPR000719; -

CC InterPro: IPR001245; -

CC InterPro: IPR003066; -

CC Pfam: PF00047; igf.3.

CC Pfam: PF00069; PKINASE.1.

CC PRINTS: PR00109; TYRKINASE.

CC PROSITE: PS00107; PROTEIN_KINASE.

CC PROSITE: PS00109; PROTEIN_KINASE_TYR.1.

CC PROSITE: PS00011; PROTEIN_KINASE_DOM.1.

CC Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;

CC Transferrase; Phosphorylation; Transmembrane; Signal;

CC Immunoglobulin domain; Alternative splicing.

CC CHAIN 1 21

CC DOMAIN 22 821

CC TRANSMEM 378 398

CC DOMAIN 399 821

CC DOMAIN 55 114

CC DOMAIN 172 238

CC DOMAIN 271 349

CC NP_BIND 481 495

CC BINDING 517 517

CC ACT_SITE 626 626

CC MOD_RES 657 657

CC DISULFID 62 107

CC DISULFID 178 231

CC DISULFID 278 342

CC CARBOHYD 83 83

CC CARBOHYD 123 123

CC FIBROBLAST GROWTH FACTOR RECEPTOR 2.

CC EXTRACELLULAR (POTENTIAL).

CC POTENTIAL.

CC CYTOPLASMIC (POTENTIAL).

CC IG-LIKE DOMAIN.

CC IG-LIKE DOMAIN.

CC IG-LIKE DOMAIN.

CC IG-LIKE DOMAIN.

CC IG-LIKE DOMAIN.

CC ATP (BY SIMILARITY).

CC ATP (BY SIMILARITY).

CC BY SIMILARITY.

CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

CC POTENTIAL.

CC POTENTIAL.

CC N-LINKED (GLCNAC. . .) (POTENTIAL).

CC N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CAROHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VARSPLIC 37 37 E -> G (IN SHORT ISOFORM)
FT VARSPLIC 38 37 MISSING (IN SHORT ISOFORM)
FT VARSPLIC 314 361 AAGVTTDKREIVLYIRNVTEDAGEYTCAGNSIGISFHS
FT CONFLICT 169 169 AMLVLP -> HSGINSSNAEVLALFNTEEDAGEYTCVS
FT CONFLICT 187 187 NYIGANQSAWLTVLPKQO (IN SHORT ISOFORM)
SQ SEQUENCE 821 AA; 91983 MW; FQDB28AD61F4414 CRC64;

Query Match 4.3%; Score 102.5; DB 1; Length 821;
Best Local Similarity 24.0%; Pred. No. 5.7;
Matches 78; Conservative 35; Mismatches 109; Indels 103; Gaps 18;

QY 182 CFSNIFIAKL-----EAVEYAHISRLQADL--NCLSDL-----LKSISQMKYIN 225
DB 491 CGQYVMAEAVGIDKDKPKENAVTA-VKMLKQDATEKDLSDLVSEMKMKIGKRNIN 549
QY 226 L--CGQDFPLASNFELVSELKLNANLLETVPNSKLEFTHHEIRRVPEYVKLP 282
DB 550 LIGACTQDPL--YVIVYASGKINREYLRRAPPGME--YSDINRPEEQMTF- 600
QY 283 IRTNISKEAPPHNIOFVSAYFV-----LSQAFYVIFNNSIVQDFMKSQDTSP 334
DB 601 -----KDLVSCYQIARMEYIASOKCTHRDLAARNVLTENNWKIADFGIARDINN 654
QY 335 DEHFMTLIRVPGIPGEISRSADVSDLSQKTRLVKWNYYEGFPYPSGSHLSVCYIG 394
DB 655 DYKKTNTGRLP-----YKMAPEALFDRVYT--HQSDVWSFG 690
QY 395 -----AAELRLKDGHWANKRDSKVDPIILKCLAEKLEBQRD-W 435
DB 691 VLMWEITLIGSPYPGIPVELEFKLEGH---RMDKPTN---CTNE-LYMMRDCW 740
QY 436 ITLPSEK-LF-----MDRLTTTS 453
DB 741 HAVPSQAPTFKQVLEDRLITLT 765

RESULT 11
YCF2_PINTH STANDARD; PRT; 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 244.6 KDa PROTEIN (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakaugui T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugita M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC - FUNCTION: NOT YET KNOWN.
CC - SIMILARITY: BELONGS THE YCF2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D17510; BAA04460.1; -.
DR InterPro: IPR001939; -.
DR Pfam: PF00004; AAA; 1.
KW chloroplast, Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 2054;
Best Local Similarity 21.5%; Pred. No. 23;
Matches 72; Conservative 60; Mismatches 98; Indels 105; Gaps 20;

QY 15 KVFILFLTW-LLSLKLINVRLEFPQKDIYVEY-SLSTSPFYRNKYTHVKDEYREVN 72
DB 864 QTFVLIDYLVRSFNLRLNP--FVREKRLSSIEISITPTLKEDI-----VNEEKN 914
QY 73 -CSGYDEPPEIGKSLERRODIIDLEDDVYAMTSQCDIYQTLRGYAKRLVS-----K 126
DB 915 FCGPFFKRSDEENNPQCFKR-----GFSNNVGLIQ-REYQDOLLSSEMSNK 962
QY 127 EKSEFPYASLVHKDALMRELIHAIYNQHNITCIHDKRAPPTFVAMNNIAKCSNI 186
DB 963 NEEIFPRT-----QDFVETELCKNKIYNE-----DIDGRS-----TSSNSKEBONT 1004
QY 187 FIASKLEVEYAHISRLQADLNCLSDLKSSIQ-----KYVINCGQDFPLKSNFEL 239
DB 1005 YRISQIDSI-----TSKW-----DLFKTYMPFPTSAWKYIENNL-----LDY 1043
QY 240 VSELKLGAN---METVAPNSKLEFTHHEIRRVPEYVKLPDIRNISKEAPPHN 295
DB 1044 LSEI-LHGSNPFSIIIONIK-HNILLKR-NILNELSHPLMEPIQCKLRNTLN----- 1094
QY 296 IQIFGSAFYLSQAFVYIFNNSIVQDFMWSKD 330
DB 1095 -----KFFPSNNKDFEYCKD 1112

RESULT 12
YCF2_TOBAC STANDARD; PRT; 2280 AA.
AC P09976; P09977;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 267 KDa PROTEIN (ORF 2280).
GN YCF2.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanales; Solanales; Solanales; Solanales; Solanales;
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, BRIGHT YELLOW 4;
RA Sugita M.;
RL Submitted (AUG-1986) to the EMBL/GenBank/DBJ databases.
RN [2]
RP COMPLETE GENOME.
RA Shinozaki K., Ohme M., Tanaka M., Wakaugui T., Hayashida N.,
RA Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohno C., Torizawa K., Meng B.Y., Sugita M.,
RA Deo H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA Tohoh N., Shimada H., Sugita M.;
RT "The complete nucleotide sequence of the tobacco chloroplast genome:
RT its gene organization and expression.";
RL EMBO J. 5:2043-2049(1986).
RN [3]

```

RP REVISIONS.
 RA Sugita M.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOT YET KNOWN.
 CC -1- SIMILARITY: BELONGS THE YCF2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 200044; CAAT7427.1; -
 DR EMBL; 200044; CAAT7438.1; -
 DR PIR; A05204; A05204.
 DR PIR; A05205; A05205.
 DR InterPro: IPR001939; -
 DR Pfam; PF00004; AAA; 1.
 KW Chloroplast; Hypothetical protein.
 SQ SEQUENCE 2280 AA; 266812 MW; E246D5F3D902C06D CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 2280;
 Best Local Similarity 20.0%; Pred. No. 26;
 Matches 88; Conservative 67; Mismatches 169; Indels 117; Gaps 20;

QY 9 KHTLQKRVFLTLMLSLKTLNVRRLPQKDIYLVEXSLSTSPFRNRTHVDEVR 68
 DQ 761 KDTLNHTIMKYITINQYLSNLSKQKWFPLILSTRSMNDP-----DAYR 810
 QY 69 YEYVSGIYQEPLEIKSLERRDIDDDDDVAMTSDCIYQLR--GYA---QKL 123
 DQ 811 YKWSNKGNEQEHLEQSVSEKSEFYV-----VFDRLRINQYSDMSEV 854
 QY 124 VSKSEKSPFLAYSLVHKDAIMVERLIHAYNCHNYC-----YDRKAPD 170
 DQ 855 IDKDKSLKPLRFL--SKSLFLKLFELFNSLPFCVGFNPIRSEIYIELKGFN 912
 QY 171 TFKAMNNLAKCFENIFASK-----LEAVEYAHISRLADLNCSDLKSSIQKYYI 224
 DQ 913 --DOLQMLLESIGLYVHLKMKKFFLLDDHDISQKSKFPLNGTISFPLKNIP--KWI 969
 QY 225 NLQGDPELKN-----FELVSEIKLKGANMLETVK--PNSKLEEFYHHEL 271
 DQ 970 D-----SFTRNRKRKSPDNDSYFSMIFH---DQDNMLNPKVPEHRSLSISPYKANRL 1021
 QY 272 R-----RVPEYVKLPRTNLSKEAPPHNIOIFVOSAFVLSQAFVKYIN 317
 DQ 1022 RFLNNPHFCEYWNTRPEFVEYKARINNSDFTYGGQPLNI-LFINKLFSLCVGKKHAF- 1079
 QY 318 NSIYQDFEAMSKDYSPDEHFWATLIRVGPISRSADVDSDLSQKTRLVKWNVEGF 377
 DQ 1080 -----MGKDTISPIESQVSNIF--IPNDFQSGDET-----YNLKSF 1115
 QY 378 FYPSCTGSHLSVCITYGAEL 398
 DQ 1116 HPEPSDPEVRR-AIYSIADI 1135

RESULT 13
 BACB_BACLI STANDARD; PRT: 2607 AA.
 AC 068007;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BACTIRACIN SYNTHETASE 2 (BA2) [INCLUDES: ATP-DEPENDENT LYSINE
 DE ADENYLAASE (LYSA) (LYSINE ACTIVASE); ATP-DEPENDENT D-ORNITHINE
 DE ADENYLAASE (LYSA) (LYSINE ACTIVASE); ORNITHINE RACEMASE
 DE (EC 5.1.1.12)].
 GN BACB.

OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 10716;
 RX MEDLINE=98089193; PubMed=9427658;
 RA Konz D., Kleus A., Schoergerdorfer K., Marahiel M.A.;
 RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
 RT 10716: molecular characterization of three multi-modular peptide
 RT synthetases." 1997-937(1997).
 RL Chem. Biol. 4:927-937(1997).
 CC -1- FUNCTION: ACTIVATES TWO AMINO ACIDS AND INCORPORATE A D-ORNITHINE
 CC FROM ITS SECOND ACTIVE SITE INTO BACTIRACIN.
 CC -1- CATALYTIC ACTIVITY: L-ORNITHINE - D-ORNITHINE.
 CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETHEINE
 CC (POTENTIAL).
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
 CC ANTIBIOTIC BACTIRACIN.
 CC -1- SUBUNIT: LARGE MULTISUBUNIT COMPLEX OF BA1, BA2 AND BA3.
 CC -1- DOMAIN: CONSISTS OF TWO MODULES WITH A C-TERMINAL EPIMERIZATION
 CC DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO THE PEPTIDE
 CC PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR
 CC SUBSTRATE ADENYLAATION, THIOALATION, CONDENSATION (NOT FOR THE
 CC INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND N
 CC METHYLATION (OPTIONAL).
 CC -1- MISCELLANEOUS: BACTIRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
 CC DODECAPETIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
 CC ABUNDANT IS BACTIRACIN A, A BRANCHED CYCLIC DODECAPETIDE. IT
 CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (LIE-CYS-LEU-D-
 CC GLU-TLE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
 CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-TLE-D-PHE-
 CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
 CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
 CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
 CC PHE-9, AND ASP-11).
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF007865; AAC06347.1; -
 DR InterPro: IPR000255; -
 DR InterPro: IPR000873; -
 DR InterPro: IPR001242; -
 DR Pfam; PF000501; AMP-binding; 2.
 DR Pfam; PF000668; DUF4; 3.
 DR Pfam; PF000550; pp-binding; 2.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
 DR PROSITE: PS00455; AMP-BINDING; 2.
 DR PROSITE: PS50075; ACP-DOMAIN; 2.
 KW Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
 KW Multifunctional enzyme; Repeat.
 FT REPEAT 535 1090
 FT REPEAT 1547 2141
 FT DOMAIN 1021 1037
 FT DOMAIN 2064 2104
 FT DOMAIN 1051 1051
 FT BINDING 2094 2094
 FT BINDING 2094 2094
 SQ SEQUENCE 2607 AA; 297474 MW; FF654FAC5B8BBA6F CRC64;

Query Match 4.2%; Score 100.5; DB 1; Length 2607;
 Best Local Similarity 19.8%; Pred. No. 37;
 Matches 85; Conservative 71; Mismatches 128; Indels 145; Gaps 25;

```

Query Match Similarity      4.2%; Score 99.5; DB 1; Length 666;
Best Local Similarity      22.0%; Pred. No. 7.1;
Matches 75; Conservative 50; Mismatches 111; Indels 105; Gaps 18;

QY 46 VEYSLSTSPFVNNRTHYHDEYREVENOSGIEODEPLEIGKSLER----- 92
DB 292 VEVNISYDVE---RYRQFADEFRDYIM---IKERRQIIMOSGDRIRFRFRPMSLSTLIK 345
QY 93 RDIDLED-----DDVY-----AATSCDIYQTLRGAAQ 121
DB 346 KDTDSLEEDILAHIDNARKNSKYSIEDVARIISSEFLNCFVARRTLSDIDITKTI--MYL 403
QY 122 KLVSEKSESPFLAYSLVYHKDAIMVERLILAHYIHOHNITCYIHYDRK-----APDTFKVA 175
DB 404 KIV-KDMKSCALITLSAI---KGIMVTDIDINIVLSK-----LEHHNNVRYKYLTSVENKELA 455
QY 176 MNNLAKCSNFIPIASKLEAVEYAHISLRQADIMC-----LSDLKSIQMKYIYNICG 228
DB 456 VCNCRSICS-----LEFRREKLSYRCDLRTDGDGLDRLYDITRALHSGKINQNIIG 505
QY 229 ODF--PLKSNFELVSELKKNLANLEUVKPPNSLSEFTYHHELRPRYEVYKPIFTN 286
DB 506 ORCWGPDLEMLFENEKKKKKLN--NUMETIKIS---DMLVYGHST-----EXTLIPIIDS 554
QY 287 ISKEAPPHNIQILEVGSAYFVLSQAFVK-YIFNNSIVQDFFA 326
DB 555 LSEKLSVDYDMS-----VLNDQYAKLIVFENTITIEYITIA 587

RESULT 15
CEK3_CHICK
ID CEK3_CHICK STANDARD; PRT; 823 AA.
AC P18461;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TYROSINE KINASE RECEPTOR CEK3 PRECURSOR (EC 2.7.1.112).
GN CEK3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90332672; PubMed=2165604;
RA Pasquale E.B.;
RT "A distinctive family of embryonic protein-tyrosine kinase
RT receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:5812-5816(1990).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE TROBOLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M35196; AAA48665.1; -.
DR PIR, B35963; B35963.
DR HSSP, P11362; IFGI.
DR InterPro: IPR000719; -.
DR InterPro: IPR001245; -.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; Ig_3.
DR Pfam: PF00069; PKinase_1.
DR PRINTS, PR00109; TYRKINASE.
DR PROSITE, PS00107; PROTEIN_KINASE_ATP_1.

```


DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW Transferase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 823 TYROSINE KINASE RECEPTOR CEK3.
FT DOMAIN 24 379 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 380 400 POTENTIAL.
FT DOMAIN 401 823 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 117 IG-LIKE DOMAIN.
FT DOMAIN 174 240 IG-LIKE DOMAIN.
FT DOMAIN 273 351 IG-LIKE DOMAIN.
FT DOMAIN 135 145 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 483 772 ASB/GLU-RICH (HIGHLY ACIDIC).
FT NP_BIND 489 497 PROTEIN KINASE.
FT BINDING 519 519 ATP (BY SIMILARITY).
FT ACT_SITE 628 628 ATP (BY SIMILARITY).
FT MOD_RES 659 659 BY SIMILARITY.
FT DISULFID 65 110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT DISULFID 181 233 POTENTIAL.
FT DISULFID 280 344 POTENTIAL.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 823 AA; 92299 MW; 42B3C64EA02FD43 CRC64;

Query Match 4.2%; Score 99.5; DB 1; Length 823;
Best Local Similarity 24.0%; Pred. No. 9.4;
Matches 78; Conservative 34; Mismatches 110; Indels 103; Gaps 18;

QY 182 CFSNFIASKL-----EAVEAHISRLOADL--NCISDL-----LKSSIQMKYVIN 225
DB 493 CFGGVNAEAVGIDKRPKEAVTVA-VKMLKDATEKOLSDLVSEMEMKMKIGKHNIIN 551
QY 226 L---CGODFLKSNFELVSEIKINGANMLETVKPNKSLERFTYHHELRVPYEVKLP 282
DB 552 LLGACTQDDGL--YVIVEYASKGNLREYLARARPCME-----YSPDINRVPEQMTF- 602
QY 283 IRTNISKAEPPHNIQIVGSAHYV-----LSQAFVKYIFNNSTIVQDFFANSKDTYSP 334
DB 603 -----KDLVSCYQLARGMEYLASQKCIHRDLAARNVLTENNVMKRIADFGIARDINNI 656
QY 335 DEHFATLIRPGIPGELISRSADVDLSQKTRLVKWNVEGFFYPSCTGSHLRVCITYG 394
DB 657 DYVKKTNGRLP-----VKMAPEALEDRYVT--HOSDVMSTG 692
QY 395 -----AAELRWLIDGHWFANKEDSKVDPILIKCLAEKLEEQORD-W 435
DB 693 VLWMEIFTLGSSPYPGIPVELEFLKEGH-----RMDKPPAN-----CTNE-LYMMHDCW 742
QY 436 ITPSEK-LF-----MDRNLTTTS 453
DB 743 QAVPSQRPTRFQVLVEDDRIILTLTT 767

Search completed: August 2, 2001, 14:50:18
Job time: 298 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2001, 00:11:58 ; Search time 113.87 Seconds

(without alignments)
7510.336 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362

Sequence: 1 atgaagataatcaatgta.....atctactacacacatcatga 1362

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.GeneSeq_0601.*

- 1: /SIDS8/gcgdata/geneSeq/geneSeq/NA1980.DAT.*
- 2: /SIDS8/gcgdata/geneSeq/geneSeq/NA1981.DAT.*
- 3: /SIDS8/gcgdata/geneSeq/geneSeq/NA1982.DAT.*
- 4: /SIDS8/gcgdata/geneSeq/geneSeq/NA1983.DAT.*
- 5: /SIDS8/gcgdata/geneSeq/geneSeq/NA1984.DAT.*
- 6: /SIDS8/gcgdata/geneSeq/geneSeq/NA1985.DAT.*
- 7: /SIDS8/gcgdata/geneSeq/geneSeq/NA1986.DAT.*
- 8: /SIDS8/gcgdata/geneSeq/geneSeq/NA1987.DAT.*
- 9: /SIDS8/gcgdata/geneSeq/geneSeq/NA1988.DAT.*
- 10: /SIDS8/gcgdata/geneSeq/geneSeq/NA1989.DAT.*
- 11: /SIDS8/gcgdata/geneSeq/geneSeq/NA1990.DAT.*
- 12: /SIDS8/gcgdata/geneSeq/geneSeq/NA1991.DAT.*
- 13: /SIDS8/gcgdata/geneSeq/geneSeq/NA1992.DAT.*
- 14: /SIDS8/gcgdata/geneSeq/geneSeq/NA1993.DAT.*
- 15: /SIDS8/gcgdata/geneSeq/geneSeq/NA1994.DAT.*
- 16: /SIDS8/gcgdata/geneSeq/geneSeq/NA1995.DAT.*
- 17: /SIDS8/gcgdata/geneSeq/geneSeq/NA1996.DAT.*
- 18: /SIDS8/gcgdata/geneSeq/geneSeq/NA1997.DAT.*
- 19: /SIDS8/gcgdata/geneSeq/geneSeq/NA1998.DAT.*
- 20: /SIDS8/gcgdata/geneSeq/geneSeq/NA1999.DAT.*
- 21: /SIDS8/gcgdata/geneSeq/geneSeq/NA2000.DAT.*
- 22: /SIDS8/gcgdata/geneSeq/geneSeq/NA2001.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	22 AAS00045	Human DNA encoding
2	191.8	14.1	1221	21 AAS06575	Core 2 beta-1,6-N-
3	191.8	14.1	1317	21 AAS06569	DNA encoding a cor
4	191.8	14.1	2108	21 AAS06570	A core 2 beta-1,6-
5	191.8	14.1	2128	22 AAC62134	Nucleotide sequenc
6	191.8	14.1	2229	21 AAC09109	Human pancreatic c
7	191.8	14.1	2319	21 AAS048623	Human C2/4GNT CDNA
8	185.2	13.6	1807	16 AAG08201	I-branchning enzyme
9	185.2	13.6	1807	17 AAT16201	Beta-1,6-N-acetylgl
10	185.2	13.6	1807	19 AAV30006	Full length CDNA s
11	185.2	13.6	1807	19 AAV16000	Human beta-1,6-N-a

12	183	13.4	186	21 AAC19209	Human secreted pro
13	175.2	12.9	2105	15 AAG61559	CDNA sequence of h
14	174.4	12.8	5010	19 AAX24042	Rat DH1 CDNA. Rat
15	174.4	12.8	5010	21 AAC65468	Diabetic rat heart
16	172	12.6	2204	19 AAX24043	Human core 2Gnt DN
17	172	12.6	2204	21 AAC65469	Human heart core 2
18	160.8	11.8	2102	18 AAT80112	CDNA encoding a re
19	121.2	8.9	378	19 AAV30005	Oligonucleotide D1
20	112.4	8.3	936	22 AAF58252	Oligonucleotide D1
21	112.4	8.3	936	22 AAF58254	Oligonucleotide D1
22	112.4	8.3	936	22 AAF58257	Oligonucleotide D1
23	112.4	8.3	936	22 AAF58259	Oligonucleotide D1
24	112.4	8.3	936	22 AAF58262	Oligonucleotide D1
25	112.4	8.3	936	22 AAF58265	Oligonucleotide D1
26	110.6	8.1	936	22 AAF58252	Oligonucleotide D1
27	110.6	8.1	936	22 AAF58254	Oligonucleotide D1
28	110.6	8.1	936	22 AAF58257	Oligonucleotide D1
29	110.6	8.1	936	22 AAF58259	Oligonucleotide D1
30	110.6	8.1	936	22 AAF58262	Oligonucleotide D1
31	110.6	8.1	938	22 AAF58255	Oligonucleotide D1
32	92.2	5.8	378	17 AAT16202	C2GNT catalytic do
33	80.8	5.9	997	19 AAF59800	Human secreted pro
34	66.2	4.9	300	21 AAT01261	Human colon cancer
35	53.4	3.9	192	22 AAC62139	Murine beta-1,6-N-
36	49.4	3.6	777	19 AAF59681	Human secreted pro
37	47.4	3.5	307	21 AAC01854	Human secreted pro
38	46.2	3.4	6171	19 AAF52170	Streptococcus pneu
39	41.8	3.1	244	22 AAF58238	Oligonucleotide D1
40	40.6	3.0	244	22 AAF58238	Oligonucleotide D1
41	40.6	3.0	7430	19 AAV31250	E. coli y96 pathog
42	35.8	2.6	814	20 AAX22268	Human secreted pro
43	35.2	2.6	99	17 AAT16203	C2GNT C-terminal r
44	35.2	2.6	322	21 AAT79284	Human lung tumour
45	35	2.6	733	20 AAT16673	Human gene express

ALIGNMENTS

RESULT 1	
AAS00045	standard; DNA: 1362 BP.
XX	
AC	AAS00045;
XX	
DT	14-MAY-2001 (first entry)
XX	
DE	Human DNA encoding C2GNT3.
XX	
KW	Human; C2GNT3; Thymus-related disorder; cancer; tumour; adenoma;
KW	UPP-GICNAC; Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
KW	malignant melanoma; breast cancer; cervical cancer; hypocoactivity;
KW	hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KW	leukemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
KW	acquired immunodeficiency syndrome; sepsis; wound healing; infection; ds.
OS	Homo sapiens.
XX	
PH	Key
FT	CDS
FT	1..1362
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= "C2GNT3"
FT	complement (114..138)
FT	/*tag= b
FT	/label= "Binding site for PCR primer TSHC100"
FT	1367..1362
FT	/*tag= c
FT	/label= "Binding site for PCR primer TSHC121"
PN	W0200114535-A2.
XX	
PD	01-MAR-2001.
XX	

PF 24-AUG-2000; 2000MO-DK00469.
 XX
 PR 24-AUG-1999; 99US-0150488.
 XX
 PA (SCHW/) SCHWIENTEK T.
 PA (CLAU/) CLAUSEN H.
 XX
 PI Schlientek T, Clausen H;
 XX
 DR MPI: 2001-226515/23.
 DR P-FSDB: AAU00037.
 XX
 PT New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
 PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
 PT thymus-related disorders, cancers, tumours, immunosuppression
 XX
 PS Claim 5; Fig 1; 97pp; English.

The sequence encodes Human UDP-N-acetyl-glucosamine:Galactose-beta1,
 3-N-acetylglucosamine-alpha-R beta1-6 N-acetylglucosaminyltransferase
 (UDP-GlcNAc:Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
 C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
 preparation of compositions for treating a conditions mediated by C2GnT3,
 particularly a thymus-related disorder. C2GnT3, nucleic acids
 encoding it and antibodies against it may also be used for in vitro
 purposes related to scientific research, DNA synthesis and manufacture of
 vectors, in the prognostic and diagnostic evaluation of conditions
 associated with altered expression or activity of C2GnT3 or conditions
 requiring modulation of C2GnT3, as well as in monitoring conditions by
 detecting and localising the DNA and protein. Disorders such as tumours
 (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 breast or cervix), hypoparathyroidism, hyperactivity, atrophy, enlargement of
 thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression,
 acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
 sepsis, wound healing, acute and chronic infection, cell-mediated or
 humoral immunity, or TNF/TN2 imbalance, may be treated using these protein
 or nucleic acid. The antibodies may be used to screen potential
 therapeutic compounds to determine their effects on a conditions such as
 thymus-related disorder or cancer, to determine the level of C2GnT3
 expression in cells genetically engineered to produce C2GnT3, or to
 detect and quantify polypeptides in a sample to determine their role in a
 particular cellular events or pathological states and to diagnose and
 treat such pathological states.

Sequence 1362 BP; 410 A; 260 C; 279 G; 413 T; 0 other;

Query Match 100.0%; Score 1362; DB 22; Length 1362;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatattcaaatgttattttaacatacctacagcagaagtttcatcctgtt 60
 Db 1 atgaagatattcaaatgttattttaacatacctacagcagaagtttcatcctgtt 60
 QY 61 ttaacccatgtgctgctctcttctttaaagctttaaagtgtgagaagcctcttcgcaa 120
 Db 61 ttaacccatgtgctgctctcttctttaaagctttaaagtgtgagaagcctcttcgcaa 120
 QY 121 aaagacattacttggtgagtaactccttaagtaacctgaccttttgaagaagaagttac 180
 Db 121 aaagacattacttggtgagtaactccttaagtaacctgaccttttgaagaagaagttac 180
 QY 181 actcatgttaagatgaagtcagatgaagtttaactgttctgggtatcatgaagaagag 240
 Db 181 actcatgttaagatgaagtcagatgaagtttaactgttctgggtatcatgaagaagag 240
 QY 241 cctttggaattggaagaagtcctggaataagaagaaggagacatcattgagtgagat 300
 Db 241 cctttggaattggaagaagtcctggaataagaagaaggagacatcattgagtgagat 300
 QY 301 gatgatgttggtggaatgacagatgattgtgacattatcagactctaagaagttatgct 360
 Db 301 gatgatgttggtggaatgacagatgattgtgacattatcagactctaagaagttatgct 360

Db 301 gatgatgttggtggaatgacagatgattgtgacattatcagactctaagaagttatgct 360
 QY 361 caaaagcttgctccaagaaggagaaagcttcccaatgactatcttggttccac 420
 Db 361 caaaagcttgctccaagaaggagaaagcttcccaatgactatcttggttccac 420
 QY 421 aaagatgcaattatgtgttaaaagcctatccatgcatatatacaacagacatttac 480
 Db 421 aaagatgcaattatgtgttaaaagcctatccatgcatatatacaacagacatttac 480
 QY 481 tgcattcatatgatactgtaaggacactgataccttcaaaagtgcattgaacaattagct 540
 Db 481 tgcattcatatgatactgtaaggacactgataccttcaaaagtgcattgaacaattagct 540
 QY 541 aagtgcttcccaatttttcatgttcttccaaattagagagcgtgtgaatcccaat 600
 Db 541 aagtgcttcccaatttttcatgttcttccaaattagagagcgtgtgaatcccaat 600
 QY 541 aagtgcttcccaatttttcatgttcttccaaattagagagcgtgtgaatcccaat 600
 Db 541 aagtgcttcccaatttttcatgttcttccaaattagagagcgtgtgaatcccaat 600
 QY 601 tccagatccagctgatttaatttctgttcgacacttgaagtcctcaatccagttg 660
 Db 601 tccagatccagctgatttaatttctgttcgacacttgaagtcctcaatccagttg 660
 QY 661 aatatcttatcacttggttgaggcaagatttccctgaagtcgaattttgattggtg 720
 Db 661 aatatcttatcacttggttgaggcaagatttccctgaagtcgaattttgattggtg 720
 QY 721 tccagatgtaaaaaactcaatgaggaatatactgttgagacggtgaaaccccaacagt 780
 Db 721 tccagatgtaaaaaactcaatgaggaatatactgttgagacggtgaaaccccaacagt 780
 QY 781 aaattggaagatctcttaccatcatcagatagagtgagtgagtgagtgagtgag 840
 Db 781 aaattggaagatctcttaccatcatcagatagagtgagtgagtgagtgagtgag 840
 QY 841 ctaccataaaggacaacatctccaaaggaagaccccccataacattcaatttgg 900
 Db 841 ctaccataaaggacaacatctccaaaggaagaccccccataacattcaatttgg 900
 QY 901 ggaagtcattatttctttaaagcaagattgttaaatattttcaaacaccatc 960
 Db 901 ggaagtcattatttctttaaagcaagattgttaaatattttcaaacaccatc 960
 QY 961 gtccaagacttttctgctggttcaagaacacatactctccgatatgagactttggct 1020
 Db 961 gtccaagacttttctgctggttcaagaacacatactctccgatatgagactttggct 1020
 QY 1021 acctgattcgggttccagaataactcgtggagatttccagatccagcagaatgtgtct 1080
 Db 1021 acctgattcgggttccagaataactcgtggagatttccagatccagcagaatgtgtct 1080
 QY 1081 gatctgcaagatgaagactcgtcctgtccaagtgaattactaagaagccttttccatcc 1140
 Db 1081 gatctgcaagatgaagactcgtcctgtccaagtgaattactaagaagccttttccatcc 1140
 QY 1141 agttgactggaactcctacacttgaagcgtgtgtattatgagagcgtgcaaatgaagttg 1200
 Db 1141 agttgactggaactcctacacttgaagcgtgtgtattatgagagcgtgcaaatgaagttg 1200
 QY 1201 ctatcaagaatgagacatggttctgtaataatttgatttcaaggtgagacctatctgt 1260
 Db 1201 ctatcaagaatgagacatggttctgtaataatttgatttcaaggtgagacctatctgt 1260
 QY 1261 attaatgcttgagcagaagaagcttgaagaacagagagactgatacattgacctca 1320
 Db 1261 attaatgcttgagcagaagaagcttgaagaacagagagactgatacattgacctca 1320
 QY 1321 gaaaagtattatgagatagaagaatccactacacatactga 1362
 Db 1321 gaaaagtattatgagatagaagaatccactacacatactga 1362

RESULT 2
 AAA96575

ID AAA96575 standard; DNA; 1221 BP.
 XX
 AC AAA96575;
 XX
 DT 08-FEB-2001 (first entry)
 DE
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNA.
 XX
 KW Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory bowel disease; arteriosclerosis;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis; ss.
 XX
 OS Homo sapiens.
 XX
 PN CA2296936-A1.
 XX
 XX 03-AUG-2000.
 PD
 XX 03-FEB-2000; 2000CA-2296936.
 PF
 XX 03-FEB-1999; 99US-0118674.
 PR
 XX (GLXC-) GLYCODESIGN INC.
 XX
 XX Korczak B, Lew A;
 XX
 XX WPI; 2000-594746/57.
 DR
 XX
 XX
 PT New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 PT
 PS Claim 4; Page 54; 66pp; English.
 PS
 XX
 XX The present sequence encodes a fragment of a human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.
 CC
 XX Sequence 1221 BP; 342 A; 285 C; 295 G; 299 T; 0 other:
 XX

Query Match	Similarity	Score	191.8	DB 21	Length	1221	
Best Local	Similarity	52.5%	Pred.	No. 4.2e-41			
Matches	517	Conservative	0	Mismatches	432	Indels	36
					Gaps	3	
Qy	317	tgaccagttattgtgacatttaccagacttaagaagttatgtccaaaaagctgtgtctcaa	376				
Db	221	tcaccagagactcttgagcacttcaagctcgaagaagttaccatcacggtccccaatgagca	280				
Qy	377	aggaagagaagaagcttcccaatgactatcttggttgtgccaaagaatgcaattatgg	436				
Db	281	aaagaagaagtgagcttcccatctgcaaccctatgctgcatctcctgaagaagattgaaact	340				
Qy	437	ttgaagagcttaccatctgcataataaacagagacacatttcatgcatccattatgact	496				

[illegible]

KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis; ss.

OS Homo sapiens.

Key Location/Qualifiers
 CDS 1..1317

FT /*tag= a
 FT /product= "core 2
 FT Beta-1,6-N-acetylglucosaminyltransferase"

PN CA2296936-A1.

XX 03-AUG-2000.

PF 03-FEB-2000; 2000CA-2296936.

PR 03-FEB-1999; 99US-0118674.

PA (GLYC-) GLYCDESIGN INC.

PI Korczak B, Lew A;

DR WPI: 2000-594746/57.

DR P-PSDB; AAB18995.

XX New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.

XX Claim 4: Page 50; 66pp; English.

XX The present sequence encodes a human core 2

CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.

CC The polypeptide can be used to treat diseases and disorders, such as

CC cancer, cardiovascular disorders and inflammatory disorders including

CC asthma, rheumatoid arthritis, inflammatory bowel disease, syndrome

CC arteriosclerosis, septic shock, adult respiratory distress syndrome

CC (ARDS) and cancer. Various platelet-mediated pathologies such as

CC atherosclerosis and clotting can also be treated. The polypeptides of

CC the invention are predominantly expressed in gastrointestinal tissue

CC (stomach, colon, intestine, testis) and are elevated in cancer.

CC Gastrointestinal disorders that may be prevented or treated include

CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis

CC and ulcerative colitis. The antibodies may be used in

CC immuno-histochemical analysis, to detect the novel polypeptide and to

CC localize it to particular cells and tissues and to specific subcellular

CC locations and to quantitate the level of expression.

XX Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;

Query Match 14.1%; Score 191.8; DB: 21; Length 1317;
 Best Local Similarity 52.5%; Pred. No. 4.3e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagttattgtacattatcagactcagaaggttatgtcacaagaatgtcaca 376
 DB 317 taccagagactgtgagcttcaagctggaaggaattatatacagttcccaactgagca 376
 QY 377 aggaagagaagaagcttcccaatagctatttctgtgttcacacaagaatcgaattatg 436
 DB 377 aagaagaggttgagcttccattatgcatctactatgttattatgagaagattgaaact 436
 QY 437 ttgaagagttcatcatgataatacaacagacacattatcagctcattatgac 496
 DB 437 ttgaagagttcatcatgagctgtgtgtgtcccccacaaataacgttccatgtgagtg 496
 QY 497 gtaagagcattatcattcaaatgtgcacatgaacaattttagtaagtgttccacaata 556
 DB 497 agaagtcctccagaacttcaagaagcggttcaagaacattatcttctgtcccaatg 556

QY 557 ttctacgttcccaaatatagagctgtggaatatgcccacatttccagactcagagctg 616
 DB 557 tcttcatagcagtaagagctgttccgtgtgttattatgctctccgtgccaggtgtcagagctg 616
 QY 617 atttaattgtctgttcgaagcttctgaagcttcaatccagttgaataatgtttacact 676
 DB 617 acctcaactgcattgagaagcttctccagagctcagtcgctggaataacttccgaata 676
 QY 677 tgtgtgtggaagatttcccccgaagtcacaatttgaattgtgttcagaggttgaataaac 736
 DB 677 catgtgtgagcagcttccctcaataagacagacagagatgtgtccaggtctccaagatgt 736
 QY 737 tcaatgagcaaatatgtgtgagacgttgaaaccccccaacagtaattgaaagattca 796
 DB 737 tgaatgtgaggaattagcatgagagtgatccctctaagacacaaagaaacccgtgga 796
 QY 797 cttaacatcagtaacttaagaggtgtccttatgataatgtgaagttaccacaagaagaca 856
 DB 797 aatatactttgaggtagtagagacacattacac-----ctaacca 838
 QY 857 acatcccaagagagacccccccaataacatcagaatttgttcagagttatttg 916
 DB 839 acagaagaagagatccctcccttaataattatcattgtttacaggaatgcatgtg 898
 QY 917 tttaagtcagcaattgtttaataatattcaacaaccatcagttcaagacttttg 976
 DB 899 tggcttcccgagatttgcaccaacatgttttgaagaaccttaattccacaacactgattg 958
 QY 977 ccggtgtcaagacacatctcctcagtagagacttggcttaacttgcaggttc 1036
 DB 959 aatgtgttaaaagacactttagccagatgaaacactctcgtggccacaccttcagctgac 1018
 QY 1037 caggaatacctgtggagatttccagatccagccagatg---tgctgtatcgcagagta 1093
 DB 1019 ggtgagatgctgtgtctgttcccaaccacccaagtagacacatctgaagcatgttcta 1078
 QY 1094 agactcgcctgtcgaagtgaattactatgaaggtcttcttcaactt----- 1143
 DB 1079 ttgcacagctgtgcagatgagcagatgagaggaacatgataaggtgtcctctatg 1138
 QY 1144 -----tgtactgtatcctcaactcgaagcgtgtgtattatgagctgcagaattaggt 1198
 DB 1139 ctccctctcttgaatccacacagcggctctctcgttattggtgtgtggttgaatt 1198
 QY 1199 ggcattacaagatgtgacattgttgcataataattgtatcattaggtggaccattct 1258
 DB 1199 ggaatgttcaaaacacatcactgttggccacaagtttgaaccacaagtagatgataatg 1258
 QY 1259 tgattaaatgttggcagaagaagct 1283
 DB 1259 ctctcagtgcttagaagaatacct 1283

RESULT 4

AAA96570 standard; DNA; 2108 BP.

AAA96570;

08-FEB-2001 (first entry)

A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.

Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;

cancer; cardiovascular disorder; inflammatory disorder; asthma;

rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;

septic shock; adult respiratory distress syndrome; ARDS; cancer;

platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;

clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;

diverticulitis; ulcerative colitis; ss.

OS Homo sapiens.

CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99237 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;

Query Match 14.1%; Score 191.8; DB 21; Length 2229;

Best Local Similarity 52.5%; Pred. No. 5.4e-44; Mismatches 432; Indels 36; Gaps 3;

Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtgatgtgacattatcaccagctcaagagtgatctgctcaaaagctgtctca 376
 Db 743 taccagagactgtgagacttcaagctggaagagtgatctatcagcttccactgagca 802

QY 377 agggaggaagaaagcttcccaatagctattttgtgtgtccacaagaagtccaattg 436
 Db 803 aagaagagtgagagctccctatgcatctctatgtgtatcacaagagagattaaact 862

QY 437 ttgaagagctatcatcgtatatacaacacgaacatatttccatgcatcattgac 496
 Db 863 ttgaagagctactgagctgtgtatgtccctcagaacatattactgtctcatgtgagt 922

QY 497 gtaagagcccgatcccttcaaatgtgcatgaaacattgtcgaagtgtcttcccaata 556
 Db 923 agaagtcctccgaacattcacaagagcggtcaagaacattattctgtctcccaatt 982

QY 557 ttctatgcttcccaattagaagcggtgtgaaatagccacattccagatccagagctg 616
 Db 983 tctcatagccagtagctgtgtgtgtgtttatgacctgtgtccaggtgtgacagtg 1042

QY 617 atttaattgctgtcgaaccttctgaagcttccatccagttggaataatgttatcaact 676
 Db 1043 acctcaactgcatggaagactgtctccagagctcagtcggtgaaatactctcgaata 1102

QY 677 tgtgtgggcaagatttcccccgaagtcacaatttgaattgtgtcagaagttgaaacac 736
 Db 1103 catgtgggagagacttctctataaagacatgacagatgtgtccagctctcaagatgt 1162

QY 737 tcaatgagcaaatatgtgtgagacggtgaaaccccaacagtaaatgtgaaagatca 796
 Db 1163 tgaatggagagatagcatgagtgatcagatcctcctaagacacaaacccgctgga 1222

QY 797 cttaaccatcatgaacttgagcggtgcttgaatcatgtgaagtaccacaataaggaaca 856
 Db 1223 aatataccttggagtgatgagaagacataac-----ctaacca 1264

QY 857 aacatcccaaggaagaccccccaataacatcagatatgtgttcagtgctattttg 916
 Db 1265 acaagaagagagatcctcccttataatttaacatgtttaacaagagatcgctacatgt 1324

QY 917 tttaagtcacagcatttttaatatatttcaaacactcatcgtttaaactttttg 976
 Db 1325 tggcttcccgagatttcgcaacatgttttgaagaccccaataatcccaacactgattg 1384

QY 977 cctgtgtctaaagacacatactctctgtatgagacattttgggtactgtattcggttcc 1036
 Db 1385 aatgggttaaaagacacttatagccagatgaaacacctctgggcccaccttcagcggtgac 1444

QY 1037 caggaataccitggggagatttccagatcagccagagat---tgctatgtcagagta 1093
 Db 1445 gttgtgtcctgtgctctgttcccaaccaccagaagacatctcagacatgacttca 1504

QY 1094 agactgccttgcgaatggaatttactatgaagcttttctatccagct----- 1143
 Db 1505 ttgcagctcgtgcgaatgagggatgagggagacatgataaggtgtctctatg 1564

QY 1144 -----tgtatgtatctcccttcgaagcgttgttttatagagctgagatataagt 1198
 Db 1565 ctccctgcttgcgaatccaccagcggtctatctgcgttatatgggctgggagcttgaat 1624

QY 1199 ggttatcaagaatgagacattgttgcataaattatgataagtgagacccatct 1258
 Db 1625 ggtatgttcaaaacccatccctctgttgccacaagattgacccaagtgagtataatg 1684

QY 1259 tgattaatgtctgcgcaaaaagct 1283
 Db 1685 ctctcagttgtctagaagataact 1709

RESULT 7

AAA48623
 ID AAA48623 standard; CDNA; 2319 BP.

AC AAA48623;

DT 19-SEP-2000 (first entry)

XX Human C2/4GNT CDNA.

XX Human; C2/4GNT; UDP-N-acetylglucosamine; O-glycan biosynthesis;

KW O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 486..1812

FT /tag- a

FT /product- "C2/4GNT"

FT primer_bind 586..605

FT /tag- b

FT primer_bind 1794..1813

FT /tag- c

FT polyA_signal 2244..2249

FT /tag- d

XX WO200034449-A2.

XX 15-JUN-2000.

XX 03-DEC-1999; 99MO-DK00677.

XX 04-DEC-1998; 98DK-0001605.

XX (Clau/) CLAUSEN H.

XX Clausen H, Schwientek T;

XX WPI; 2000-423407/36.

XX P-PSDB; AAY94452.

XX New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
 XX probe for the detection of specified glucosaminyltransferase from
 XX other species and related organisms
 XX
 XX Claim 5; Fig 2; 47pp; English.
 XX
 XX The present sequence encodes human UDP-N-acetylglucosamine:
 XX N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
 XX (C2/4GNT). The protein is the third member of the family of O-glycan
 XX beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
 XX based O-glycans on to oligosaccharides, glycoproteins and

glycosphingolipids. C2/4GNT can therefore be used in the production of appropriately glycosylated glycoconjugates with particular enzymatic, immunogenic, or other biological or physical properties. The nucleotide sequence is useful as a probe for the detection of C2/4GNT from other species and related organisms and for the recombinant production of C2/4GNT polypeptide. The nucleotide sequence was identified by analysis of EST database sequence information. Oligonucleotides derived from EST clone 17865 of ATCC were used to isolate two full-length C2/4GNT clones from a human foreskin genomic PL library by 5' RACE PCR. RT-PCR was performed using Colo205 human cell line mRNA in order to produce cDNA for expression of C2/4GNT in Sf9. The control of O-glycan core assembly has been implicated in tumour progression and metastasis.

Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other;

Query Match 14.1%; Score 191.8; DB 21; Length 2319;
Best Local Similarity 52.5%; Pred. No. 5.5e-44;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtgatctgacattatcatcagactcgaaggtatgctcaaaagcttgctcaa 376
DB 812 tcaccagagactgtagagactcgaaggtatgctcaaaagcttgctcaa 871
QY 377 agggaggaagaaagctcccaatagctattcttggtgtgtccacaagaatgcaatttg 436
DB 872 aagaagaggtgagctccctatctgacactctatgctgacatcagagaagatgaaact 931
QY 437 ttgaagagcttatccatctatatacaacacacacacacacacacacacacacacacac 496
DB 932 ttgaagagcttatccatctatatacaacacacacacacacacacacacacacacacac 991
QY 497 gtaaggaacctgataacctcctcaaatgccaatgacaatgataagtgctctccata 556
DB 992 agaagtcctccagaaactctccaaagagcggtcaaaagcaatattctgtccccaatg 1051
QY 557 tttaattgctcccaaatgataagcgctgtggaatagtcacacacacacacacacacac 616
DB 1052 tcttatagccagatagctgtgtcggtgtttatgctcctgtgtccaggggtgaagctg 1111
QY 617 attaaatgtctgtcgaagctcttgaaagcttcaatccagtggaatattgtatcaact 676
DB 1112 aacctaaactgcatggaagactgtccagagctcagtgccgtggaataactctctgata 1171
QY 677 tgtgtgggcaagatttcccttgaaagtcacacacacacacacacacacacacacacac 736
DB 1172 catgtgggaaggaacttccctataagacacacacacacacacacacacacacacac 1231
QY 737 tcaatgagcaaatatgtttgagacagtgtaaaccccccaacagtaaatgtgaagaattca 796
DB 1232 tgaatgaggaagatagcagtgagtgacacacacacacacacacacacacacacacac 1291
QY 797 cttaacatcatgaacttgaagcgtgtgcttgaatatagtgaagctacacataagagca 856
DB 1292 aatacacttctgaggtagtgagagacacattacac-----ctaacca 1333
QY 857 acatctcaaggaagacaccccccaacacacacacacacacacacacacacacacac 916
DB 1334 acaagaagaagatcctccctataatatttaactatgttacaagggatcgtfacattg 1393
QY 917 tttaagcagaacattgtttaatattttcaacaacacacacacacacacacacacacacac 976
DB 1394 tggctcccgagattctgccaacatgttttgaagaacacacacacacacacacacacac 1453
QY 977 cctgtctaaagacacacactcctctgtagacacacacacacacacacacacacacacac 1036
DB 1454 aatgggtaaagacacactatagccagtagaacacacacacacacacacacacacacacac 1513
QY 1037 caggaataactcgggagatttccagatcagccagagatg---tgtctgattcgaagatga 1093
DB 1514 ggtgattcctcgtgtctgttcccaacacacacacacacacacacacacacacacacac 1573
QY 1094 agactcgcctgtcaagtggaattactatgaagcttttctatccagct----- 1143

DB 1574 tgcacagctgtgtcgaagtgtagcaggtatcagtagagagacatgataaggtctctctatg 1633
QY 1144 -----tgtactgtatcacaacttcgaagcgtgtgtattatgagctcgaagaaatgct 1198
DB 1634 ctccctgccttgaaactccaccagcggcgtctatctgcttattatggtcgtggagctgaaat 1693
QY 1199 ggcattcaaaagatgtagcttggttcttaataattgattcctaaggttgacacactatc 1258
DB 1694 ggaatgcttaaaaac 1753
QY 1259 tgattaatgcttggcagaagaagct 1283
DB 1754 cctctcagtgcttagaagaataact 1778

RESULT 8
AA089201
ID AA089201 standard; cDNA; 1807 BP.

AA089201;

23-SEP-1995 (first entry)

I-branching enzyme cDNA.

Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IGNT;

Inflammation; antiinflammatory; tumor; hypersensitivity; anemia;

transgenic animal; I antigen; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 255..1457

W09507020-A.

16-MAR-1995.

09-SEP-1993; 93WO-US08476.

09-SEP-1993; 93WO-US08476.

(LJOL-) LA JOLLA CANCER RES FOUND.

Bierhuizen MFA, Fukuda M;

WPI; 1995-123182/16.

P-PSDB; AAR71932.

Nucleic acid encoding I-branching enzyme - used to develop prods.

for treating e.g. inflammatory responses, tumours or

hypersensitivity reactions

Claim 1; Fig.6; 56pp; English.

Poly-A RNA isolated from human PA-1 teratocarcinoma cells was.

reverse-transcribed and inserted into expression vector pCDNA1.

Plasmid DNA was used to transfect CHO-PV-10u cells. Transfected

cells were screened using human anti-I antigen antibodies and goat

anti-human IgM. Plasmid DNA was analyzed to obtain DNA (given in

AA089201) encoding I-branching enzyme (AAR71932).

Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other;

Query Match 13.6%; Score 185.2; DB 16; Length 1807;

Best Local Similarity 52.1%; Pred. No. 3.5e-42; Indels 24; Gaps 3;

Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

327 ttgtgacattatcagactcgaaggtatgctcaaaagctgtgtctcaaaagagagaga 386

XX	22-APR-1996	(first entry)
DE	Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme CDNA.	
RX	Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme; IGNT;	
KW	glycosyltransferase; blood group; I antigen; polylactosaminoglycan;	
KM	ds.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	
FT	CDS	
XX	Location/Qualifiers	
XX	255..1457	
XX	/tag= a	
PN	US5484590-A.	
XX		
PD	16-JAN-1996.	
XX		
PF	09-SEP-1993; 93US-0118906.	
XX		
PR	09-SEP-1993; 93US-0118906.	
XX		
PA	(LJOL-) LA JOLLA CANCER RES FOUND.	
PI	Bierhuizen MFA, Fukuda M;	
DR	MP1; 1996-087019/09.	
DR	P-PSDB; AAR92474.	
PT	Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme	
PT	- used to develop prods. for the study, detection and treatment of	
PT	pathological conditions involving the enzyme.	
XX		
PS	Example 4; Column 29-32; 29pp; English.	
CC	A CDNA insert (AAT16201) in pCDNA1-IGNT codes for human beta-1,6-N-	
CC	acetylgalactosaminyltransferase, I-branching enzyme (IGNT) (AAR92474),	
CC	an enzyme that converts blood group I antigen to I antigen. The	
CC	CDNA was isolated from a cDNA library pred. from human PA-I	
CC	tartocarcinoma cells in vector pCDNA1 following transfection into	
CC	CHO-Py-leu cells and screening with human anti-I antibodies. The	
CC	CDNA can be used to regulate the expression of human IGNT or to	
CC	modify its biological function, to produce soluble or membrane-bound	
CC	forms of IGNT in host cells, to breed transgenic animals, and to	
CC	design antisense oligonucleotides.	
SQ	Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other:	
	Query Match 13.6%; Score 185.2; DB 17; Length 1807;	
	Best Local Similarity 52.1%; Pred. No. 3,5e+42;	
	Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps	
	3	
OY	327 ttgtgcatattacgaacttaagagtatgctcaaaagctgtctcaagaaggagaa 386	
DB	467 ttccaagatatcttgaccacagcactataccacaagcccccttctaaggaagaag 526	
OY	387 aaagttcccgaatgctatcttgtgtggcccaagaagagcaattatggttgaaggt 446	
DB	527 tgaccttccttgcataataatgltcacccatcacatttgacaccttgcagaagct 586	
OY	447 tatccatgcataataaacagacacaaatttatcctgatcatatgatcgtaaagc 506	
DB	587 ctccaagcatttaactatgcccaaaaataatctactgtgtcatgtgataaaaaaac 646	
OY	507 tgaacctcaaaagttgcatgaacaatttaagctaaagtgtcttccaattttcaatgc 566	
DB	647 aactgaatttaagaatgcgttagagcactaatgaagtgtcttcccaaagcgttcgcgc 706	
OY	567 ttccaattgaagcgttggataatgcccatttcagagctccagagctatattaatg 626	
DB	707 ttccaagaatggaacccgtttctcatgtaggagatctcagagctccagctgaacctgaactg 766	

QY 1167 cgtgtattatgagctgcagaaatlaagctggtctatccagaagatgagcattggttgc 1226
 DB 1283 tattgtatctatgaaagaggaacttaagctggtggttaattccacagccttggc 1342
 QY 1227 taataattgattcgaagctgaccctatcttgattaaatgctgacgaagaacttga 1286
 DB 1343 taagaattgagcttaataccacccttaactgagatgctcagaactgagcactcg 1402
 QY 1287 agaa 1290
 DB 1403 cga 1406

RESULT 12

AA061559
 ID AAC19209 standard; cDNA; 186 BP.

AC AAC19209;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 23284.

KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 23284; 71bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 186 BP; 54 A; 37 C; 38 G; 57 T; 0 other;

Query Match

Best Local Similarity 13.4%; Score 183; DB 21; Length 186;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 tttaacctatgctgctctcttctttaaagcttcaatgtagacgaactcttcgca 119
 |||

DB 1 tttaacctatgctgctctcttctttaaagcttcaatgtagacgaactcttcgca 60
 QY 120 aaaagacattactggttgcagtaacctccctaagactgccttcttgaagaacagata 179
 DB 61 aaaagacattactggttgcagtaacctccctaagactgccttcttgaagaacagata 120
 QY 180 cactcatgttaagagcaagctgaagatgaagttactgttcgggtctcagaacagga 239
 DB 121 cactcatgttaagagcaagctgaagatgaagttactgttcgggtctcagaacagga 180
 QY 240 gcc 242
 DB 181 gcc 183

RESULT 13

AA061559
 ID AA061559 standard; cDNA; 2105 BP.

AC AA061559;

DT 07-OCT-1994 (first entry)

DE cDNA sequence of human core 2 beta 1-6

DE N-acetylglucosaminyltransferase (C2GNT or core 1-6 AGT).

KM C2GNT; 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;
 O-glycan; ss.

OS Homo sapiens.

PN EP590747-A.

PD 06-APR-1994.

PF 29-SEP-1993; 93EP-0250268.

PR 01-OCT-1992; 92US-0955041.

XX (LJOL-) LA JOLLA CANCER RES FOUND.

PI Blerhuizen MFA, Fukuda M;

DR WPI; 1994-111195/14.

DR P-PSDB; AAR51386.

XX New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are
 PT used for the study of the effect of variant O-glycan(s) on
 PT cell-cell interactions, partic. in cancers

PS Disclosure; Page 20-22; 34pp; English.

CC C2GNT, or an active fragment thereof, catalyses the formation of
 CC critical branches in O-glycans. cDNAs encoding various
 CC glycosyltransferases can be isolated by transient expression of cDNA
 CC in recipient cells, e.g. COS-1. COS-1 cells were transfected
 CC with a cDNA library, pCDSR alpha-2FL1, constructed from poly(A)+ RNA
 CC of activated T lymphocytes which express the C2GNT. Transfected
 CC cells were selected using MAB T305, which identifies a
 CC hexasaccharide on leukosialin. Leukosialin CD43 is an acceptor
 CC molecule for C2GNT activity. A plasmid, pCDSR alpha-Leu,
 CC which directed expression of the T305 antigen was identified. The
 CC cDNA insert was isolated and sequenced. The cDNA encoded the
 CC acceptor molecule leukosialin CD43.

XX	Sequence	2105 BP;	614 A;	402 C;	500 G;	589 T;	0 other
50							

Query Match	12.9%;	Score 175.2;	DB 15;	Length 2105;
Best Local Similarity	51.9%;	Pred. No. 2.5e-39;		
Matches 517;	Conservative	0;	Mismatches 443;	Indels 36;
				Gaps 4

QY	300	tgtatgattgtgttgcaatctgacagctgattgtgacattttctgacattcttgacattgaaggtgtacgc	359
Db	489	tgcagctatataaactgacccgagctgacgtgtctctcttccatcaagagagcccaataatat	548
QY	360	tcaaaagctgtctcctcaaggagagaaaagctcccaatagccatctctgtgtgtccca	419
Db	549	tgtgaagaccctctagttaaaagaaagagcggaagttcccaatagatattctatgtgtgtcca	608
QY	420	caaaagatgcattatggtttgaagggttttccatctgatatatacaaccagcaaatatta	479
Db	609	tccaaagattggaatgcttgcacagctgtctgtagtccatctatccctcagaatttcta	668
QY	480	ctgcattccattatgattcgttaaggcacccatgataccttcaaaagtgtccatgaaacaattgac	539
Db	669	ttcgattcatgttgacacaacaacccagagattccctattgaagctcagttgagtcagtcgc	728
QY	540	taagtgcttcccaatttttcaattttctcccaattagaagggctgtggaaatggcccaat	599
Db	729	tctccgttttaagtaagctctgtctgtgcccagccgagctggagctgggttttagcactgtg	788
QY	600	tccagaactccagagctgtatttaaatgtctgttcgagaccttccgaaagcttccatccagt	659
Db	789	gagccgaggttccagctctgacctcactgcatgaaagattctctatgcaatgaaacaactg	848
QY	660	gaaatagtataactgtgtgtggagaaattttccctcgaaagtcaaattttgattgtgt	719
Db	849	gaatgaactctatcaactctgtgtgtatgtgatttcccaatlaaaccaacctgaattgtc	908
QY	720	gtcagaattgaaaaaacctcaattgtagcaataatgtgtgagagagctgtaaacccccaacag	779
Db	909	caggaagctcaagtglttaatgttgagagaaacaacctggaacggagagatgcacccca	968
QY	780	taaatgtgaaagtttaactataccatcatatgaacttgaagcggtgcttataatagttaa	839
Db	969	taaaggaagaaaggtctgagaaagacggatga-----gagcgttaaaggaaagctgac	1019
QY	840	gtcaccaataagagacaacatctcccaaggagaacaccccccataacatcagataattgt	899
Db	1020	-----aaacacagggagctgtcaaaatgtctctccacatcgaaacacctctcttc	1070
QY	900	tggagagctattttgttttaagtcaaacattgtttaatatatttccaaacactcat	959
Db	1071	tggagagctactctcgtctgtcgtatgagatgatgtgtgtgtatgtatcactacgaatgaa	1130
QY	960	catctcaagactttcttgcgcgcgtctaaagaacatatctctctctga tgaagcatttggc	1019
Db	1131	aatccaaagttgatgagtgagtgggcacagaacacatcagcctgtagtattctctggc	1190
QY	1020	tactctgattccggttccaggaatatactctggggagat-----tccagatcagcccgagatgt	1076
Db	1191	caccatccaaagattctcctaagttcccggggtcaacctccctgcgcgcataagtaagtatct	1250
QY	1077	gtctgattcagagatgaagctgcgctcttgatcgaatggaatlaactatgaagcgtttt---	1133
Db	1251	atctgacatctgaagcagcttgccaggttggttcaagtgtagtaactttgggggtgagtttc	1310
QY	1134	-----ctaccacagttgtatcctgagatctcaactctgaaagcgtgtattattarg	1181
Db	1311	caaggtgtctccctacacgcgcctgcgcatgagatcctatgctgcgcataagttgtcatttcg	1370
QY	1182	agctgcgaatattaagtggtcttcaagaatgacatgcatgtgttgcataaataattgtatc	1241
Db	1371	agctggtgacttgtaacgtgagtgctgtgcgaaacacacactgttgcataaattgtcagct	1430
QY	1242	taaggtgacacctatcttgatttaaatgcttggcaga	1277

Db 1431 ggaatgtgacctcttggccatccagtggttggatga 1466

```
RESULT 14
'AAx24042
ID AAx24042 standard; cDNA; 5010 BP
```

AC	AAx24042.
XX	
DE	28-JUN-1999 (first entry)
XX	
DE	Rat DH1 cDNA.
XX	
DH1	rat; screening; treatment; prevention; cardiomyopathy; inhibitor;
KW	diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate
KW	UDP-GlcNAc:alpha1-3galNAc-alpha; transgenic animal; germ line;
KW	beta-1,6-N-acetylglucosaminyl-transferase; ss.
XX	

OS Ratus sp

EH	Key	Location/Qualifiers
EH	Key	Location/Qualifiers

$$/*tag = a$$

XX
XX

CAZ100381-A
XX

PD 02-APR-1996.
XX

PF	02-001-1996;	96CA-218698/
XX		

PR 02-001-1996; 96CA-218698/.
XX
XX

PA (MOUN) MOUNT SINAI HOSPITAL CORP.
XX

PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE, XY

DR WPI; 1998-399608/35.
DR R-PSDB; 20003043

XX
XX

PT associated with diabetes and hyperglycaemia - comprises reacting

PT in presence of test substance

PS Example 1; Fig 3A; 35pp; English

CC This invention describes a method for screening for a substance that
CC may be used to prevent or treat cardiomyopathy associated with diabetes
CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
CC (UDP-GlcNAc:aldehyde-3GalNAc-alpha beta-1,6-N-acetylglucosaminyl-
CC transferase) with an acceptor substrate and a sugar nucleotide donor in
CC the presence of a test substance under conditions whereby the core 2
CC GlcNAc-T produces a reaction product, determining the amount of reaction
CC product, and comparing the amount of reaction product with the amount
CC obtained in the absence of the test substance, where lower amounts of
CC reaction product in the presence of the test substance indicate that the
CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
CC methods for preventing or treating cardiomyopathy associated with
CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
CC germ cells and somatic cells all contain a DNA construct introduced into
CC the animal or an ancestor of the animal at an embryonic stage, where
CC incorporation of the DNA construct into the germ line of the animal
CC causes the animal to develop cardiomyopathy similar to that associated
CC with diabetes mellitus and hyperglycaemia. This sequence encodes the rat
CC DH1 protein which is used to describe the method of the invention.
XX
SQ Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match	12.8%;	Score 174.4;	DB 19;	Length 5010;
Best Local Similarity	51.8%;	Pred. No. 6e-39;		

Matches 518; Conservative 0; Mismatches 446; Indels 36; Gaps 4;

```

QY 302 atgagtgtggaacatgacagtgatgtgacattatcagactcaagaggtatgctc 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1073 atgactatataaacatgacccggagctgcgcctcctcattacagacagcaatatatta 1132
QY 362 aaaagcttgcctcaaaaggagaaaagctcccaatagcctcttctgtgtccaca 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1133 tggagcccttactaaagaaaggttgccttcccaattgatatccatagttgttcacc 1192
QY 422 aagatgcaattatggttgaaggcttaccatctatataccacagcccaatttact 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1193 ataatgattgacatgcttgacagcctcctgagccatctataagccagaaatttact 1252
QY 482 gcatccattatgactgtaaggacccctgataccttcaaatgtccatgaacaattagcta 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 gcatccacgtgacagaaagacaggaatccctttagccggcgagcaggtatgcat 1312
QY 542 agtgccttcccaatttcttctgcttcccaattagaagctgtgagatagcccaattt 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1313 cctgcttgaataagctcttctgtgcccagcagcttgagaaagtctgtatccgctctgga 1372
QY 602 ccagactccagctgacttaaatgtctgtcgcacctctgtaagttcattccactgga 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1373 gtccgggttaagctgacactcaactgcgtgaagacctgtacagaatgaatgcaactgga 1432
QY 662 aatatgtatcaactgtgtgtggcgaagatttccctggaagttcaaatlttgattgtgt 721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1433 agtactgtacaaactctgtgtgtatgtattccctatataaacacccggaattgttca 1492
QY 722 cagagtgtgaaaacaaactcaatgagagcaataatgttgaagcgtgtaaaccccaagta 781
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1493 ggaagctcaagctcctctacacgggaaaacgcctggaactggaagatgacctcccaaca 1552
QY 782 aattgaaagatccacttaccatcagaaactgaacgggtgtccttgaatatgtgaagc 841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1553 aggaagagaggttgaa-----aaacagatcacagctgttgagcgggaaagc 1597
QY 842 taccaataagagacaacatctccaaagaaagcacccccatacattcagatatgtgtg 901
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1598 t-----gaaaacaactgtgagtgtcaaggcgagcctccactcaaatctctcttctcag 1654
QY 902 caggtgtctattctgttttaagtcgaagcattgttaatatatttcaaacactcatcg 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1655 gcagcgccctattcgtgtcactaggaatatgtagctgtgtgctggaataaataaata 1714
QY 962 tcaagaacttttgcctgtgtctaaagaacacatcctcctgtatgagcactttgtgcta 1021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1715 tcaaaaagttcatgaaatgtggcacaagacacatacagcccgatgtctcctctggtgcca 1774
QY 1022 ccttgattcgggttccagaaataccttggggagatttccagatcagccagagtg--tgt 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1775 ccacccaagaagatccctgaaagctcgtgtctctccctcaagacccaagtaagttactgt 1834
QY 1079 ctgactcggagatgaagactcgcctgtgtcaagtggaattactatgaaggttttcc----- 1134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1835 ctgacaagaaatgcgtgcacgaagttgtcaagtggaataacttcgaagcgagatgttcca 1894
QY 1135 -----tatccaagtgtactgtgactcacttcgaaacggtgtatattatgag 1183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1895 atggcgcccttatccacgcgtgagtgagtgatcattgtgcctctgtgtcttggag 1954
QY 1184 ctgcagaattaaagtggtctatcaaaagatggaactgtgttgctataaatttgattcca 1243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1955 ttgtgtcattgagctgagtgctgtcgcaaaacacacttttttccataagtttgacatgg 2014
QY 1244 aggtggaacctatctgtatlaaagtcttggcagaagaact 1283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2015 atgtggaatcccttgcctccagtggttggaaagaacatct 2054

```

RESULT 15
AAC65468

```

ID AAC65468 standard; cDNA: 5010 BP.
XX
AC AAC65468;
XX
DF 12-FEB-2001 (first entry)
XX
DE Diabetic rat heart core 2 GlcNAc-T coding sequence.
XX
KW Rat; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
KW UDP-GlcNAc:Galbeta1-3GalNAcalpha1-6
KW N-acetylglucosaminyltransferase; ss.
XX
OS Rattus sp.
XX
PN US613578-A.
XX
PD 17-OCT-2000.
XX
PF 02-OCT-1997; 97US-0943058.
XX
PR 02-OCT-1996; 96US-0046876.
XX
PA (KING/) KING G L.
PA (NISH/) NISHIO Y.
PA (KOYA/) KOYA D.
PA (DENN/) DENNIS J W.
PA (WARRE/) WARREN C E.
XX
PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
XX
WPI: 2000-678642/66.
DR P-PSDB; AAB30297.
XX
PT Preventing or treating cardiomyopathy associated with diabetes mellitus
PT and hyperglycemia, comprises administering a substance that inhibits
PT core 2 N-acetylglucosamine-T activity -
XX
PS Example 1; Fig 3A; 21pp; English.
XX
CC The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc
CC alpha1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
CC and coding sequences. The enzyme is associated with cardiomyopathy in
CC diabetes and hyperglycemia sufferers. The invention also provides
CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
CC in treatment.
XX
SQ Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

```

Query Match 12.8%; Score 174.4; DB 21; Length 5010;
Best Local Similarity 51.8%; Pred. No. 6e-39;
Matches 518; Conservative 0; Mismatches 446; Indels 36; Gaps 4;

```

QY 302 atgagtgtggaacatgacagtgatgtgacattatcagactcaagaggtatgctc 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1073 atgactatataaacatgacccggagctgcgcctcctcattacagacagcaatatatta 1132
QY 362 aaaagcttgcctcaaaaggagaaaagctcccaatagcctcttctgtgtccaca 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1133 tggagcccttactaaagaaaggttgccttcccaattgatatccatagttgttcacc 1192
QY 422 aagatgcaattatggttgaaggcttaccatctatataccacagcccaatttact 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1193 ataatgattgacatgcttgacagcctcctgagccatctataagccagaaatttact 1252
QY 482 gcatccattatgactgtaaggacccctgataccttcaaatgtccatgaacaattagcta 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 gcatccacgtgacagaaagacaggaatccctttagccggcgagcaggtatgcat 1312
QY 542 agtgccttcccaatttcttctgcttcccaattagaagctgtgagatagcccaattt 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1313 cctgcttgaataatgtcttctgtgcccagcttgagaaagtgtgtatccgctctgga 1372

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 22:43:33 ; Search time 1367.35 Seconds
(without alignments)
9415.868 Million cell updates/sec

Title: US-09-645-192-1
Perfect score: 1362
Sequence: 1 atgaagatttcaaaagtta.....atctactaccacatcatga 1362

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	# Query Match	Length	DB	ID	Description
1	437.8	32.1	821	153	BG3434801	BG3434801 602507322
C 2	408.4	30.0	410	223	AQ005888	AQ005888 CIT-HSP-2
C 3	393.8	28.9	592	245	AQ9003048	HS_2175_B
C 4	350.2	25.8	610	235	AZ6030357	AZ6030357 IM0422E09
C 5	289	21.2	292	224	AC0091453	AC0091453 HS_3016_B
C 6	253	18.6	928	220	CNS00244G	AL188089 Tetradon
C 7	221	16.2	973	222	CNS0558A	AL129563 Tetradon
8	187	13.7	1367	192	AK0199324	AK0199324 Mus muscu
9	174.2	12.8	1862	192	AK008234	AK008234 Mus muscu
C 10	170.8	12.5	860	206	AL576150	AL576150 AL576150
C 11	163.2	12.0	970	221	CNS0401M	AL5707831 Tetradon
C 12	157.6	11.6	756	220	CNS010WZ	AL168380 Tetradon
C 13	148.8	10.9	562	162	BE005007	MRO-BN011
C 14	148.8	10.9	612	145	BE292814	601105179
C 15	139.4	10.2	567	242	AZ388491	AZ388491 IM0148G03
C 16	133	9.8	172	165	BE234997	142220 MA
C 17	133	9.8	655	521	AM842632	MR2-CN003
C 18	123.8	9.8	655	24	A1722768	A1722768 f631c11_Y
19	127.8	9.4	1050	141	BE867668	BE867668 601443206
20	124.8	9.2	869	106	AL555400	AL555400 AL555400
21	121.6	8.9	695	143	BF036111	BF036111 601457742
22	118.2	8.7	731	154	BG468611	BG468611 602510243
23	116.2	8.5	842	141	BE869192	601445191
24	115.8	8.5	554	5	AA307800	AA307800 EST178656
25	113.8	8.4	671	113	AM233232	f129612_Y
26	110.8	8.1	788	21	A1538293	A1538293 u195f11_Y
C 27	107	7.9	428	228	AQ437509	AQ437509 HS_5133_B
C 28	102	7.5	348	141	BE863271	BE863271 UI-M-BHO
C 29	101.2	7.4	482	148	BF416888	BF416888 UI-R-CNO
C 30	99.6	7.3	542	115	AM388627	AM388627 MR2-ST013
31	97	7.1	633	16	A1133042	A1133042 HA1630_Hu
C 32	96.6	7.1	398	239	AZ149281	AZ149281 SP_00035_B
C 33	95.8	7.0	622	229	A1747955	A1747955 u104047_Y
34	93.4	6.9	605	229	AZ194803	AZ194803 SP_1028_B
35	92.4	6.8	555	139	BE71246	BE71246 188355 MA
36	90.6	6.7	443	169	BF755013	BF755013 QVO-CTO58
C 37	88.6	6.5	438	199	BE721377	BE721377 188533 MA
C 38	88.2	6.5	449	131	AM842590	AM842590 MR2-CN003
C 39	87	6.4	543	144	BE077138	BE077138 226944 MA
C 40	86.2	6.3	447	223	AZ150403	AZ150403 SP_0039_B
41	86	6.3	801	133	BG385709	BG385709 602454133
42	84.2	5.2	569	113	AM937414	AM937414 RC0-BN005
C 43	79.8	5.9	574	115	AM368539	AM368539 MR2-ST012
C 44	78.6	5.7	866	120	CNS0267T	AL183998 Tetradon
C 45	77.2	5.8	866	123	BG365598	BG365598 602454847

ALIGNMENTS

QY	DB	1042	121	1102	DB	181
Query Match	Best Local Similarity	32.1%	Score 437.8; DB 153; Length 821;			
Matches 439; Conservative		99.5%; Pred No. 3,2e-107;				
		0; Mismatches 2; Indels 0; Gaps 0				
QY 922	agtcgaacattctgttaataataatttcacaacatccatcgctgttccaagacttttgcctcg	961				
DB 1	AGTCGAACATTTGTAAATATATATTTCACACACATCCATCATTCATCAAGACTTTTGTGCTGG	60				
QY 982	tctaaagacacataactctctcgtatgtagcacttttggcttacctcgattcgggttcacgga	1041				
DB 61	TCTAAAGACACATACCTCTCTGATGAGACACTTTTGGGCTTACTCTTATTCGGGTTCCAGCA	120				
QY 1042	ataacctggggagatttccagatataagcccgagatgtgtctgatactgcagagtaagaactcgc	1101				
DB 121	ATACCTGGGAGATTTCACAGATACGACCCAGAGATGTGTGATCTGCAGACGTAAGACATCGC	180				
QY 1102	cttctgaagtggaattactatgaagacttttctatcccaagtgtgaacgagatccacactt	1161				
DB 181	CTTGTCAAGTGGAATTACTATGAAGGCTTTTCTATCCCAAGTTGACTGAGATCTACACTT	240				

RESULT	1			
LOCUS	BG434801			
DEFINITION	BG434801	821 bp	mRNA	EST
	602507322F1	NIH_MGC_79	Homo sapiens	CDNA clone IMAGE:4604669 5',
				mRNA sequence.
ACCESSION	BG434801			
VERSION	BG434801.1	GI:13341307		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
REFERENCE	1 (bases 1 to 821)			
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Published (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D.			

RESULT	2
LOCUS	A000588/c
DEFINITION	A000588 410 bp DNA GSS 27-JUN-1998
	CIT-HSP-2288B17.TF CIT-HSP Homo sapiens genomic clone 2288B17, DNA

sequence.
 ACCESSION AQ005888
 VERSION AQ005888.1 GI:3083333
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 410)
 Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H.,
 Simon,M. and Venter,J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building (1998)
 JOURNAL Unpublished (1998)
 COMMENT Other-GSS: CIT-HSP-2288B17.TR
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: M13-21;
 Class: BAC ends.

FEATURES
 Source
 1.410
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="GDB:7149764"
 /db_xref="taxon:9606"
 /clone_1lb="CIT-HSP"
 /sex="Male"
 /note="Vector: pBelobAC11; site_1: HindIII; site_2:
 HindIII"

BASE COUNT 135 a 87 c 81 g 107 t
 ORIGIN

Query Match 30.0%; Score 408.4; DB 223; Length 410;
 Best Local Similarity 99.8%; Pred. No. 2.2e-99;
 Matches 409; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 869 aagacaccccccaatacattcagatatttggcagtgctatttggtttaagtcgaag 928
 |||||||
 DB 410 AAGCACCACCCCAATACATTCAGATATATGTTGCACTGCTATTTGTTTAAAGTCAAG 351
 |||||||

QY 929 cattgttaataatattttaaacaactcactgcgtcaagaacttttggcctgtctaaag 988
 |||||||
 DB 350 CATTGTTAAATATATTTTCAACAGCTCATCGTTCAGACTTTTGGCTGGCTTAAAG 231
 |||||||

QY 989 acacatactcctcgtatgagcacttttggcctactcgtatcgggttcagaataacctg 1048
 |||||||
 DB 290 AACATACTCTCTGATGAGAGCACTTTGGGCTACTGATTCGGGTTCCAGAAATACCTG 231
 |||||||

QY 1049 ggggagatttcagatcggccagagatgtgtctgtctgcagagaaagacccgctgtca 1108
 |||||||
 DB 230 GGGGAGATTTCAGATCAGCCCAAGATGCTGATCTGCAGAGTAAGACCTGGCTTGTCA 171
 |||||||

QY 1109 agtgaattactaagaagcctttctctatcccaattgtaactgagctccaccctcgaagcg 1168
 |||||||
 DB 170 AGTGAATTAATGAAGGCTTTTCTATCCAGTGTACTGATGATCTCACCTTGAAGCG 111
 |||||||

QY 1169 tgtgtatttaagagctgcagaataagtggtgtcttaacaagaatgagcattgtgtcta 1228
 |||||||
 DB 110 TGTGATTATGAGAGCTGCAGAAATTAAGTGGCTTATCAAAAGATGACATTGGTTCCTA 51
 |||||||

QY 1229 ataattgattcctaagtggaacctatctgattaaatgcttgcagaa 1278
 |||||||

DB 50 ATAATTGATCTAAGGTGACCCCTATCTGATTAATGCTTGGCAGAA 1

RESULT 3
 AQ000348/c 612 bp DNA GSS 10-NOV-1999
 LOCUS HS_2175_B1.A12.T7 CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone plate-2175 Col-23 Row-B, DNA sequence.
 ACCESSION AQ000348
 VERSION AQ000348.1 GI:6356538
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 612)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 JOURNAL Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2175 row: B column: 23
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 612.

FEATURES
 Source
 1..612
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC clones in
 E-Coli DH10B"

BASE COUNT 180 a 141 c 121 g 165 t 5 others
 ORIGIN

Query Match 28.9%; Score 393.2; DB 235; Length 612;
 Best Local Similarity 91.6%; Pred. No. 3.1e-95;
 Matches 427; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 839 aactcccaataaggaagcactcctcaaggaagaccccccaatacattcagatattg 898
 |||||||
 DB 610 AGCTTCCAAATAGATGACAGCTTCTCCAAAGAGCACCCGCTCATACATTCAGATATGTG 551
 |||||||

QY 899 ttgggc-agtgcctatttgttttaagtcgaacttgttaatatatttccaacattcc 957
 |||||||
 DB 550 TTGGCAAGGATTAATTTTGTTCAAAGTCAGCATGTGTAATATATATTTCAACACTCC 491
 |||||||

QY 958 atggttcaagacttttttcccggtctaaagacacatacttcccgatgagcactttgg 1017
 |||||||
 DB 490 ATGTTCAAAACTTTTGCTGCTGCTGCTTAAGACACACTACTCTCATGATGAGCATCTTGG 431
 |||||||

QY 1018 gctacattgattcgggtttccagaataactcgtgggaatttccagatcagccagagatg 1077
 |||||||
 DB 430 GTTACTGATTCGGGCTTCAGAAATACCTGGGAGATTTCCAGATTCAGCCAGAGATGG 371
 |||||||

QY 1078 tctgacttcagagtaagactcgtcctgtcgaagtggaattactatagaagctttttcat 1137
 |||||||
 DB 370 TCTGATCTCGAGATGAAGCTCGCTTGTTCAGTGAATTAATTAAGGCTTTTCTCAT 311
 |||||||

Oy	1138	cccgattgtaactggaatctcaaccttcgaagcgtgfgtatttttgtagctgcagaattaag	1197	
Db	310	CCCGATTGTACTGGATCTCACCTTCGAAGCCTGCTGATTATTGAGCGCGCAAAATAAG	251	
Oy	1198	tggctataccaagaatgacacattggttgtctaataatttgatcttaaggvtgaccctatc	1237	
Db	250	TGGCTATTCAAAGAATGGACATGGTTGCTATTAATAATTGATTCTAAGTGACCATAATC	191	
Oy	1258	ttgatcataagtctgcgcagaaagccttbaagaacaagcagaagact	1303	
Db	190	TTGATTATATGCTTGGCGGAAAAGCTTCGCATGCCGCGCAGTCCGCT	145	
RESULT	4	AZ603057	GSS	13-DEC-2000
LOCUS		AZ603057	590 bp	DNA
DEFINITION		IM0422E09F Mouse 10kb plasmid UUGCLM library	Mus musculus genomic clone UUGCLM0422E09 F, DNA sequence.	
ACCESSION		AZ603057		
VERSION		AZ603057.1	GI:11725247	
KEYWORDS		GSS.		
SOURCE		house mouse.		
ORGANISM		Mus musculus		
REFERENCE		Maniatis; Metzger; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
AUTHORS		Dunn,D., Aoyagi,A., Barber,M., Bascorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.		
TITLE		Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts		
JOURNAL		Unpublished (2000)		
COMMENT		Contact: Robert B. Weiss University of Utah Genome Center Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112 USR		
FEATURES		Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0..00 Plate: 0422 row: E column: 09 Seq primer: CGTGTAAACGACGCCAGT Class: plasmid ends High quality sequence stop: 550. Location/Qualifiers 1..590		

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U081M0422809"
 /clone_id="Mouse 10kb plasmid U081M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD29v; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/nanres/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-treated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The ligated DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD29 (g11473111491AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells

BASE COUNT	and selected for ampicillin resistance."
157 a	147 c
138 g	147 t
1 others	

Query Match	25.8%;	Score 350.8;	DB 246;	Length 590;
Best Local Similarity	79.6%;	Pred. No. 8.4e-84;		
Matches 440;	Conservative 0;	Mismatches 108;	Indels 5;	Gaps 2

QY	1	atgaagatattcaaatlttaatttaaacacacccttaagaaagattttcaaccgttt	60
Db	38	ATGAAGATATTCAGATGTGGCTTTAAATACACTCTCCAGACAGAAACTCTTCAACCTC	97
QY	61	ttaacccctatggcgcgcctcttfttaaaotcttaaatgt---agagacctcttcg	117
Db	98	TTAAACCCTGTGGCTGTCTCTCTTTAAGTCTCTAAATGTGGGACGGCTCCCTTCCCT	157
QY	118	caaaaagacattacttggttgtagtactccctaagatcactcgcctcttglylaaagacaga	177
Db	158	CAAAAGACATTACTTACTTGTGATTACTCCCTAAGTACATCACCATTTTGGAGGAACAG	217
QY	178	tactactatgttaaggatgaagtcaggtagtaagaatttaactgcttcgggtactataagacag	237
Db	218	TTCCCCGAGTCGGGGATGACGCCAGGGACACACTTAACTGCTCTGGGGGCTTACGAGCAC	277
QY	238	gaagcctttggaatatggaagaagctctgaaataaagaagaagacatcatgtacttgag	297
Db	278	GAGCCTTTGGAAATGCGCAGAGCTCTGAAATCAGAAAGAGGAGCATCATCGACTTGGAG	337
QY	298	gatgtatgatgttbggcacatgaccagatgattgtacatatatcaagctctaaagtgat	357
Db	338	GAGGTGATGTGTGTGGCATGACAGTACGTACGTATTCAGACCCATTAAGGACGTAC	397
QY	358	gattcaaaagcttgtctcaaaagaggaanaaaagcttcccaatagactatctcttggtgtgc	417
Db	398	CATGAAAGCTGTTTCAAGGAGGAAGAGACTTCCCATAGCCTATTGCTGTGGCTGC	457
QY	418	gcacaaagtgcacattatggttga--agagcttatccatgctataataaccagacacata	475
Db	458	CACAAAGATGCCATTAATGTGTGGAGCGGGTGTATCCAGCATATTACCAACACACACACC	517
QY	476	tttactgcacatcatatgatcgtlaagcaccctgtaaccttaaaagtgcctgaacaact	535
Db	518	TTTACTGCATCATTAATGACCTGAAGTACCCGGACAGCTCAAAAGCTCCCATGAACACACC	577
QY	536	tagctaaatgacct	548
Db	578	TAGCTAAGTACTT	590

[illegible]

QY 919 ttaagtcagcattgttcaatataattccacaatccatcgltcaagcttlttggc 978
 Db 96 CTGTCGGCGAGCTTGTGGCGMAAGTGAACGAGCGAGCTGGCCAGAGTCTGCGCC 37
 QY 979 tggcttaagacacatactcttcctgagagc 1009
 Db 36 TGGTACGCTGAACCTACTCGCCGAGCAGC 6

RESULT 7
 LOCUS CNS05BAA/ 973 bp DNA GSS 26-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone
 039H15 of library A from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL329563.1 GI:8223185
 VERSION AL329563
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 973)
 Roest-Crollius, H., Jalllon, O., Dasilva, C., Fizes, C., Fisher, C.,
 Bouneau, L., Billaud, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 973)
 Roest-Crollius, H., Jalllon, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 973)
 Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 AUTHORS This sequence is a single read and was generated as part of a large-
 scale clone-and-sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source
 1..973
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_id="A"
 /note="Genoscope sequence ID: C0NA039CD08A1-end: T3"
 BASE COUNT 240 a 234 c 261 g 231 t 7 others
 ORIGIN

Query Match 16.2%; Score 221; DB 222; Length 973;
 Best Local Similarity 59.7%; Pred. No. 9.7e-49;
 Matches 371; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 181 accatcgttaagatgaagtcagatgaatgaatcgttcgtggtatcattgaacagag 240
 Db 628 AATGGTTCAGGCGCTACACAGCAAGTATGTTACTGTCCCGCATATACGAATGGAC 569
 QY 241 ccttggaaattggaagaagtcgtgaaataagaagaagagacatcgaacttgagat 300
 Db 568 CCGATGAGAGTGGGGAAGTCTTGATCAATCCGAGGAACCTCTCCGCGGAGTCGAGC 509
 QY 301 gatgagtggtggaagcagtgatgtgacattatcagacctaagagtgatct 360
 Db 508 GAAAGCTGACCAACTGACCTCTGACTGCAGAGTGTGTGAGGGCCAGAGTTACGAC 449

QY 361 caaaagctgtctcaaggaggaagaaagcttcccaatagcctatcttggttgcac 420
 Db 448 AAGAGTGGCTCTCGGAAATGAGAGAGAGCTTCCCTGGCTAGCAATGTTGTGAC 389
 QY 421 aaagatgcaattatggttgaaggcttaccatgcatatgaacagacaaattac 480
 Db 388 AATATATGCTGATGATGGAGAGCTCTCAGAGGCGGTACTCCCAATATACATC 329
 QY 481 tgcattcattatgacgtlaaggacccgtalaccctcaaaattgcatgacaattagct 540
 Db 328 TGCCTCCACTAGCACTGAAGTCCCTACAGTTATCTCGCCATAGAGGCGCTGACC 269
 QY 541 aagtgcttcccaatttccatgttcccaattgaagcgttggaatgcccacat 600
 Db 268 CGCTGTCTGCCCAAGCTCTTCATCGCGTCAAGCGGAGAGTGTTCATACGGGGCTTC 209
 QY 601 tccagactccagagctgatttaattgtctgagacccctcgaagtcacatcagtg 660
 Db 208 AGCCGGCTGAAGCCGACCTCAACTGCTGTGGACCTTCTGAGTCAAGAGTGG 149
 QY 661 aaatagttatcaactgtgtggaagatttccctcgaagtcacatcagtg 720
 Db 148 AAGTAGGTATCAACCTCTGCGGACAGATTTCCTCAGGTCACATGAGCTGTG 89
 QY 721 tcaagattgaaaaactcaatgagacaaatgttgagaagtgtaaaccccaacagt 780
 Db 88 TCAGAGCTGAAGAACTAAACGGGGCTAATGTAGAGAGCGGCGGACCAAGAGTAT 29
 QY 781 aaattggaagattcattac 801
 Db 28 AAGAGCAGAGATTCACTTTC 8

RESULT 8
 AK019924
 LOCUS AK019924 1367 bp mRNA HTC 08-FEB-2001
 DEFINITION Mus musculus adult male pituitary gland cDNA, RIKEN full-length
 enriched library, clone:530430K10, full insert sequence.
 ACCESSION AK019924
 VERSION AK019924.1 GI:12860326
 KEYWORDS CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) adult male pituitary gland cDNA to
 mRNA, clone_id:RIKEN full-length enriched mouse cDNA library
 clone:530430K10.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)
 Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Methods Enzymol. 303, 19-44 (1999)
 REFERENCE 2 (sites)
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL 20499374
 MEDLINE
 REFERENCE
 AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T.,
 Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T.,
 Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
 Matsunaga, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.
 RIKEN Integrated Sequence Analysis (RISA) system--384-format
 sequencing pipeline with 384 multichannel sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL 20530913
 MEDLINE
 REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and


```

QY 1198 tggctatcaagaatggaattgttcttaataattgattcttaaggtggaacctatc 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1317 TGGTACTCTCAGAACCATCACTCTGGCCACAAAGTTTGACCCAAAGTGATGATAAT 1376
QY 1258 ttgattaaatgcttggagaanaa 1280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1377 GTCTCTCAGCTGTTAGAGAGATA 1399

RESULT 10
AL576150 860 bp mRNA EST 16-FEB-2001
LOCUS AL576150 LTI_NFL006.PL2 Homo sapiens cDNA clone CSOD1072YD06 3
DEFINITION prime, mRNA sequence.
ACCESSION AL576150
VERSION AL576150.1 GI:12938008
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 860)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
    source
        1..860
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CSOD1072YD06"
            /clone_lib="LTI_NFL006.PL2"
            /tissue_type="placenta"
            /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed by
            Life Technologies. Contact : Feng Liang Life Technologies,
            a division of Invitrogen 9800 Medical Center Drive
            Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
            Email : fliang@lifestech.com URL :
            http://fulllength.invitrogen.com
        158 c 219 g 246 t 19 others

BASE COUNT 218 a 158 c 219 g 246 t 19 others
ORIGIN
Query Match 12.5% Score 170.8; DB 106; Length 860;
Best Local Similarity 54.4%; Pred. No. 3, 2e-35;
Matches 371; Conservative 10; Mismatches 289; Indels 12; Gaps 2;

```

```

QY 645 gtcttaatccagatgggaatatgtatcaactgtgtggtggaagatttccctgaagtc 704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 467 CTCTGAGAGTCCCTGAGAGTATTCATCTACMACCTGGGGGGAARACTTCCCTGGAAMAC 408
QY 705 aaattgaattgtgtcagagttgaaanaactcaatggagcaatatgttggagcgt 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 CACACGGGAATAGTTCAGTACTCTRAAGGATTTAAAGGAAAAATATACACCCCGGAT 348
QY 765 gaaccccccaacgtaaatgtgaagaattcaacta---ccataatgaacttagacgggt 821
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 GCTGCCTCTCTGACACGCTGTGGAGGAGCTAAATACTTCCACCAAGAACTTTAAACCA 288
QY 822 gcttaatgaatatgtgaagctaccataagaagacaacatcccaaggaagacccccca 881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 CAAAATTCCTTRCGTG-----ATTAAACACAAATAATTAATAACTCTCTCCBCA 237
QY 882 taacattcaatatattgttgcagtgcttatttgttttaagtaagaacttgttaata 941
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 TGACATGATGATTTACTTTGSCACGGCTACGTGCTCTCAAGGGACTTGTCTAATT 177
QY 942 tatttcaaacatccatcgatcaagacttttgcctggtctaaagacatactctc 1001
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 CGTCTCCAGACACGCTCGACACTTACTCTCTCGTGTCAKAGACACTACAGCC 117
QY 1002 tgatgacacttttgggtacc 1023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 GCACGACATTTCTGCTGACC 95

RESULT 11
CNS04UIM/C 970 bp DNA GSS 24-MAY-2000
LOCUS CNS04UIM/Terradon nigraviridis genome survey sequence T3 end of clone
DEFINITION 042K19 of library A from Tetradon nigraviridis, genomic survey
            sequence.
ACCESSION AL307831.1 GI:8213194
VERSION GSS; genome survey sequence.
KEYWORDS Tetradon nigraviridis.
SOURCE Tetradon nigraviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
            Tetraodontidae; Tetradon.
REFERENCE 1 (bases 1 to 970)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C.,
            Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
            Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
JOURNAL freshwater pufferfish Tetradon nigraviridis
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 970)
TITLE Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
JOURNAL Bernot, A., Fitzames, C., Wincker, P., Brodier, P., Quetier, F.,
COMMENT Saurin, W. and Weissenbach, J.
            Human gene number estimate provided by genome wide analysis using
            Tetradon nigraviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 970)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
            scale clone-and-sequencing project of the Tetradon nigraviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetradon.

FEATURES
    source
        1..970
            /organism="Tetradon nigraviridis"
            /db_xref="taxon:99883"
            /clone="042K19"
            /clone_lib="A"

```

BASE COUNT 262 a 221 c 245 g 213 t 29 others
 ORIGIN

Query Match 12.0%; Score 163.2; DB 221; Length 970;

Best Local Similarity 56.1%; Pred. No. 3,7e-33; Title
 Matches 291; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

204 gtataaagttaactgttcgtgggtatctatgaacaggagcccttggaaattggaaagagctc 263
 Db 519 GTACAGCTTATGTCTCCCGGATATAGCAATGGAACCGATGAGAGTGGGAGAGTCTTT 460
 Qy 264 ggaataaagaagaagagacatcattgacctggagagatgatgatgttggcaatgacag 323
 Db 459 GNTCATCCGGAGGAAACCTCTGCCCGAGTCCGACACAAAGCCCTGACCAAACTGACCTN 400
 Qy 324 tgattgacattatcagacttaagaaggtatgctcaaaagcttgcctcaagaaggaga 383
 Db 399 TGACTGACAGAGAGTGTGTACAGGCCAGAGGTACACAAAGAGTGCNNTTGGAAATGGA 340
 Qy 384 gaaaagctcccaatagcattcttctgtgtcccaagaagaatgcaaltatgttgaag 443
 Db 339 GCAGAGANTTCCCTCGGGTACTCATGTGTGCACANATATGCTGATGATGATGAGAG 280
 Qy 444 gttatcactgtatatacaaccagacacatatttactgcatcattatgataagc 503
 Db 279 GTCTCTCAGGCGCTGTACTCCCGCAATACATCTACTGCTCTCCACTGACCTGAACTC 220
 Qy 504 acctgataacctcaaaagtgtcgaataaattgaagtgtcctcaatatttcat 563
 Db 219 CCCCTACACAGTTCATNTGCGGCGCATAGAGGCGCTGCTGCTNTGCCAACNCTTCAT 160
 Qy 564 tgcctcaaaatgaaggcgtgtgaatatgccacattccagaccggcggatgattaa 623
 Db 159 CCGGTCCAGACGGAGAGTGTTCATACGGGGCTTCACCGCGCTGAAAGCGACCTAA 100
 Qy 624 ttgctgtcgaaccttgaagcttcaatccagctggaataatgttatacaattgtgtg 683
 Db 99 CTGCGCTGTGAGCTTCTGTAGTCAAGAGTCAAGTGAAGTACATCAACCTCTGCGG 40
 Qy 684 gaaagatttccccgaagaatcaatttgaattgtgtgc 722
 Db 39 CCAGGATTTTCCCTCAGGTCACATCGNNNTGNTNC 1

RESULT 12
 CDSOLUWZ/c 756 bp DNA GSS 12-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence PUC-Or1 end of clone
 DEFINITION 197F05 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION AL168380
 VERSION GSS: genome survey sequence.
 KEYWORDS Tetraodon nigroviridis.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 756)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizeses, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissbach, J.

TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 756)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizeses, C., Winkler, P., Brotier, P., Quetier, F.,
 Saurin, W. and Weissbach, J.

TITLE Human gene number estimate provided by genome wide analysis using

JOURNAL Tetraodon nigroviridis DNA sequence
 REFERENCE Unpublished
 3 (bases 1 to 756)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES
 source
 1..756
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="197F05"
 /clone_1ib="G"
 /note="Genoscope sequence ID : C0A6197CC03SP1-end :
 PUC-Or1"

BASE COUNT 169 a 189 c 224 g 166 t 8 others
 ORIGIN

Query Match 11.6%; Score 157.6; DB 220; Length 756;
 Best Local Similarity 61.2%; Pred. No. 1.1e-31;
 Matches 267; Conservative 2; Mismatches 166; Indels 1; Gaps 1;

Qy 848 taagagacaacatctccaaaggaagcaccaccataacattcagatattgttgcagtg 907
 Db 754 TCNAAACAGAGCTGGCCAAAGCCCGCCCGCCGCGCATCCAGKCTTCATGGCAGCG 695
 Qy 908 ctatttgtttaaagtaagcattgttaataatatttcaaacactcatcgttcaag 967
 Db 694 CTATTTCGTCTGTGTCGGCGGACTTTTGGCACACGGAACCGAGAGAGCTGCCAGCG 635
 Qy 968 acttlttgcgtgtcctaaagacacacacacacacacacacacacacacacacacacac 1027
 Db 634 ACTTCTGCGCTGTGTGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 575
 Qy 1028 ttccggttccaaagaaatccctgtggaagatttccagatcaagccagatgtgtcgtgc 1087
 Db 574 TCAGGTCGCCGGGGGCTCCCGGCGCACATCCCGCTCCCAAGCGAGTCAACGATCTGA 515
 Qy 1088 aagatgaagctcgcgtgttcaagtggaattactatgaagccttttctatcccaagtgtga 1147
 Db 514 GAAATGAAGACCGGCTGCTGTAATGAACTATCTGAGGAGAAAGCTTACCCGCGCTGCA 455
 Qy 1148 ctgatactcaccctcgaagcgtgtgtattatgaagcttgaagaattaaagtggttca 1207
 Db 454 CGGGCACACACTGTGGAGCGGTGACATCTACGCGCGCGGAGCTT-CGCTGGCTGCTG 396
 Qy 1208 aagatgaagcttgttgcataataattgatttcaagtggaagccttctgatttaaat 1267
 Db 395 GCTTCGACACTGCTGTGCGCAACAAGTTTACCCCAAGTGAAGACCGGCTGTGATCACT 336
 Qy 1268 gcttgcagaaaagct 1283
 Db 335 GTTGGAGAGAGAGCT 320

RESULT 13
 BE005007/c 526 bp mRNA EST 05-JUN-2000
 LOCUS MRO-BN0115-020300-001-a03 BN0115 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BE005007
 ACCESSION BE005007
 VERSION BE005007.1 GI:8265240
 KEYWORDS EST.

ORGANISM human.

REFERENCE 1 (bases 1 to 526)
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 2 (bases 1 to 526)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Ngai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

TITLE
JOURNAL
MEDLINE
COMMENT

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.G.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=st2-MR0-BN0115-020
300-001-a03&ts=2000-03-02&ft=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 526.
Location/Qualifiers

FEATURES
source

1. 526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BN0115"
/dev_stage="Adult"
/note="Organ: breast,normal; Vector: puc18; Site:1: Sma1;
Site:2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 126 a 125 c 136 g 139 t
ORIGIN

Query Match 10.9%; Score 148.8; DB 162; Length 526;
Best Local Similarity 60.3%; Pred. No. 2.4e-29;
Matches 246; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

345 tctaaaggttaagctcaaaagcttctcaaaaggaagaaagctcccaatagccta 404
|||||
497 TCGAAGCCACTATGTACAGAAACACTCTCTGAAGAGAGAGCGTCCCTTACTTA 438
|||||
405 tctcttggtgtccacaagaatgcaattatgtgtgaagagcttatccatataca 464
|||||
437 CACAGTACCATCCACAAAGACTTCGGCACTTTGAGCGCGCTCTTCAGGCGATTATAT 378
|||||
465 ccagacataattactgcacatataatgtatgaagcaccgtatcacctcaagtgc 524
|||||
377 GCCCCCAAAATCTACTGTGTGACCTGGATCAGAAAGCCACGATGCCCTTAAAGTGC 318
|||||
525 catgaacaattagctaaagtgtctccaaatatttcaattgtctccaattagaagctgt 564
|||||
317 AGTAAACAGATTACTACAGCTGCTCCCAAAATGCTTTCTGGCTCCAAAGAGAGTCGGT 258
|||||
585 ggaatataccacattccagactccagagctgtatcaattgtcttgccagacttcgaa 644
|||||
257 TGTATTATGGGGGATCTCCAGGCTCCAGGCTGACCTGAACCTGCTGAACACCTGTGGC 198
|||||
645 gtctcaatccagtggaataatgtatacaacttgtgtggaagaaatttccctgaagtc 704
|||||
197 CTCGAGATTCCTCGAAGTATGTCAATCAACACCCGCGGCAAGACTTCCCTGAAAG 138
|||||
705 aaatttgaattgtgtcagagttgaaaaaactcaatgagccaattat 752
|||||
137 CACACGGAATAGTTCAGTATCTGAAGGGGATTTAAAGGAAAAAATAT 90
|||||

RESULT 14
BE292814
LOCUS
DEFINITION
ACCESION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE292814 612 bp mRNA EST 13-JUL-2000
601105179P1 NIH_MGC15 Homo sapiens cDNA clone IMAGE:2988090 5',
mRNA sequence.
BE292814 GI:9175472
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 612)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: LICM76 row: m column: 19
High quality sequence stop: 606.
Location/Qualifiers

FEATURES
source

1. 612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2988090"
/clone_lib="NIH MGC.15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 181 a 137 c 140 g 154 t
ORIGIN

Query Match 10.9%; Score 148.8; DB 165; Length 612;
Best Local Similarity 54.0%; Pred. No. 2.5e-29;
Matches 340; Conservative 0; Mismatches 272; Indels 18; Gaps 1;

395 caatgacctatcttctgtgtccacaagaatgcaattatgtgtgaaggttatccatg 454
|||||
1 CTAATGCATACACTCTATMGATTCATGAGAGATTGAAGAACTTGAAGGCTACTGCGAG 60
|||||
455 catatacaaccagacaaatatttactgcattatcatatgcatgtaaggcactgtacct 514
|||||
61 CTGTGATGCCCTTCAGAACATATCTGTGTCATGTGATGAGAGAGTCCACAGAACTT 120
|||||
515 tcaaatgtccatgaacaatttagtaagtgtcttccaatatttccatttccatcaat 574
|||||
121 TCAAGAGCGGCTCAAGAACATATTTCGTGCTCCAAATGCTCTCAATAGCAGATAGC 180
|||||
575 tagaagctgtggaataatgccacaattccagaactccagagctgatttaattgtctgcg 634
|||||
181 TGTATCGGCTGTTATGCTCTCGTCCAGAGGTGCAAGCTGACCTCACTACATGAGAG 240
|||||
635 acctctgaagcttcaatccagtggaataatgtatacaacttgtgtggaagaaatttc 694
|||||
241 ACTGTGTCAGAGCTCAGAGCCGTGGAATTAATCTCTGATATCATGTGGAGCGACTTC 300
|||||
695 cccctgaagtcacatttgaattgtgtcagagttgaaaaaactcaatgagccaattatgc 754
|||||
301 CTATAAGAGCAATGACAGATGATGTCAGAGGCTCTCAAGATGTTGAATGGAGGAAATACA 360
|||||

QY	755	ttgagaagcgtttaaaaccccccaacacgtaatttggaaagatcaacttaccatcaatgaactta	814
Db	361	TGAGTACAGAGGTACCTCTTAAGCACAAGAAACCCCTGGAAATATCACTTTGAGGTAG	420
QY	815	gacgggtgccttatgaatattgtgaagctaccataaagaacaaacatctccaaggaagcac	874
Db	421	TGAGAG-----ACACATTAACACCTTAACCCAAACAAAGAAAGGATCTC	462
QY	875	cccccatcaatcagaattgttggcagtgacttatattgtctttaagccaagcatttg	934
Db	463	CCCCCTAATTAATTACTTGTGTTACAGGGATGGCTACATTGTGCTTCCCGAGATTTCG	522
QY	935	ttaaatatatttcaacacccatccatcgttccaagaacttttgcctgtgtctaaagaacat	994
Db	523	TCCAAACATGTTTGGAGAAACCTTAATCCCAACACATGATGGGTAAAGACACTT	582
QY	995	aactcctcgtatgagcactttgggactact	1024
Db	583	ATAGCCAGATGAACACTCTGGGGCAACT	612

RESULT	15
AZ388491/c	
LOCUS	AZ388491
DEFINITION	5c7 bp DNA
	1M014G03R Mouse 10kb plasmid UNGC1M library GSS
	clone UNGC1M014G03 R, DNA sequence.
	Mus musculus genomicid
	02-OCT-2000

ACCESSION	A2388491	GI:10502199
VERSION	A2388491.1	
KEYWORDS	GSS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE
Dunn, D., Aoyagi, A., Barber, M., Beacom, T., Duval, B., Hamil, C.,
1 (bases 1 to 567)

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG,
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0148 row: G column: 03
Seq primer: CACACAGCAACAGCGTATGACC
Class: plasmid ends
High quality sequence stop: 567.
Location/Qualifiers
1..567

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UBGICM0148G03"
/clone_lbp="Mouse 10kb plasmid UBGICM library"
/sex="Male"
/_id_host="E. Coli strain XL10-Gold, T1-resistant, F-"  

/mote="Vector: PWD42ny; Purified genomic DNA from M.  

musculus C57BL/6J (male) was obtained from the Jackson  

Laboratory Mouse DNA Resource  

(http://www.jax.org/resources/documents/dnares/). The DNA  

was hydrodynamically sheared by repeated passage through  

0.005 inch orifice at constant velocity. The sheared DNA  

was blunt end-repaired with T4 DNA polymerase and T4  

polynucleotide kinase. Adaptor oligonucleotides were  

ligated to the blunt ends in high molar excess. The

```

adaptor: DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g1147321149b1AR129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match	10.2%	Score 139.4	DB 242	Length 567
Best Local Similarity	54.9%	Pred. No. 8	4e-27	
Matches 275	Conservative 0	Mismatches 226	Indels 0	Gaps 0
Qy	316	atgaccagtgatctgtgacattatcagaactcctaagaaggtatctgctcaaaagctgtctca	375	
Db	561	ATGAAGCAGAGCTGTGACACACTTCAAGACCACAGAGGAGGAGTTTAAACAGGTCCACTGAGC	502	
Qy	376	aagagggagaaagcttcaccatagaccatctcttggtgtgtccacaagaagtgcattatg	435	
Db	501	AAGGAGAGGGCCAGCTTCCCATTCGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCAT	442	
Qy	436	gttgaagagcttatccatgctatataacacagacaaatattctgcatccatattatg	495	
Db	441	TTTCGAAAGGTGCTGCGAGCTGTGACCCCTCAGAAATGATCTGTCACATGAT	382	
Qy	496	cgtaagagcctgataccttcaagttgcataagcaattagctaaagtgtcttccaat	555	
Db	381	CAGAAAGCTTCAGAAACCTTTAAGAGGCAATCAGAGGCATCGTGTCACTGTTCCCAAT	322	
Qy	556	atttcattgtcttccaaattagaggctgtgtgaatatgcccacattccagaactccaggct	615	
Db	321	GTCTCAATAGGTAGTAAATGTTGTGTCAATGTGTCAATGTCTGTCACAGGTCAGGCT	262	
Qy	616	gatttaattctgtctcgagcctctctgaagctctccaatccagatgtgaatatgtataaac	675	
Db	261	GACCTAAACTGCATGGAAGAGCTTGTTTCAGAGCCCGCGCATGGAAATACCTCCTGAC	202	
Qy	676	tttgtgtggcaagaattttccccctgaagtccaattttgtaatttggtgtgcagagttgaaaaa	735	
Db	201	ACCTGTGGGACAGACTTTCATCAAAACCAATCTGTGATGTGAAGGCCCTTAAGCTA	142	
Qy	736	ctcaatgtagcaaatatcttgtagaacgttgaaccccccaaacagtaatttgaagaatc	795	
Db	141	TTGAAAGGGCAGAACACAGATGAGAGTCAGAGGTACACACTCCACATTAATAAATCCCGCTGG	82	
Qy	796	acttaccatctgaactaga	816	
Db	81	AAATTCACATGAGGTGACA	61	

Search completed: August 3, 2001, 00:31:19
Job time: 6466 sec

Fri Aug 3 09:38:07 2001

us-09-645-092-1.rst

Page 14